

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 10:02:32 ; Search time 4104 Seconds
(without alignments)
16417.679 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 1647

Sequence: 1 atgggttctaagaaataactct.....gtatggagcggtagtactag 1647

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1647	100.0	4710	1	LIU45241	U45241 Lawsonia in
2	798.8	48.5	2672	1	XBCGROEX	M86549 Amoeba prot
3	784.8	47.7	1662	1	BAP439086	AJ439086 Buchnera
4	778.4	47.3	1811	1	LPNHTPRG	M31918 Legionella
5	777	47.2	2715	1	AB057417	AB057417 Pseudotalt
6	771.2	46.8	1635	1	BGROELT	Z15160 B.bacillifo
7	771.2	46.8	1894	1	BAOB63A	M98257 Bartonella
8	770.6	46.8	1844	1	BHU78514	U78514 Bartonella
9	764.6	46.4	2241	1	PHA243594	AJ243594 Pseudotalt
10	762.6	46.3	1644	1	RQU78515	U78515 Bartonella
11	756	45.9	2016	1	LPNHTPB	M91673 Legionella
12	755	45.8	1627	1	AF014829	AF014829 Bartonell
13	753	45.7	1653	1	BAP439085	AJ439085 Buchnera
14	751.2	45.6	2770	1	LMGROELSR	X57520 L.micdadei
15	746.6	45.3	1623	6	AX607355	AX607355 Sequence
16	746.6	45.3	7265	6	AX602153	AX602153 Sequence
17	746.6	45.3	21495	1	AE014285	AE014285 Streptoco
18	746.6	45.3	89050	1	SAG766855	AL766855 Streptoco
19	744	45.2	2596	1	FTGRO	X98853 F.tularensi
20	741.8	45.0	96109	6	AX067460	AX067460 Sequence
21	741	45.0	2054	1	AF389514	AF389514 Streptoco
22	739.4	44.9	1635	1	AY256822	AY256822 Piscirick
23	738.8	44.9	1647	1	BAP439083	AJ439083 Buchnera
24	738.6	44.8	1718	1	AF325222	AF325222 Streptoco
25	732.4	44.5	5907	1	D85628	D85628 Buchnera ap
26	732.4	44.5	12782	1	AE014077	AE014077 Buchnera
27	732.4	44.5	34699	1	AF008210	AF008210 Buchnera
28	732.2	44.5	1901	6	AX416527	AX416527 Sequence
29	729	44.3	213251	6	AX413015	AX413015 Sequence
30	729	44.3	258650	1	AL596171	AL596171 Listeria
31	729	44.3	349980	6	AX417047	AX417047 Sequence
32	727	44.1	290029	1	AE017024	AE017024 Bacillus
33	725.8	44.1	11763	1	U32736	U32736 Haemophilus
34	725.8	44.1	110000	6	AR274513_05	Continuation (6 of
35	724.4	44.0	2021	1	BSP344977	AJ344977 Buchnera
36	724.4	44.0	2028	1	BSP344978	AJ344978 Buchnera
37	723.4	43.9	300169	1	AE016801	AE016801 Vibrio vu
38	722.6	43.9	2156	1	AF353323	AF353323 Listeria
39	722.6	43.9	11149	1	AE015019	AE015019 Streptoco
40	722.6	43.9	295050	1	AL591982	AL591982 Listeria
41	722.6	43.9	349980	6	AX641670	AX641670 Sequence
42	722.6	43.9	349980	6	AX641671	AX641671 Sequence
43	721.2	43.8	1647	1	AF387863	AF387863 Buchnera
44	721.2	43.8	2200	1	HEAGROSL	M91030 Haemophilus
45	719.6	43.7	1647	1	AF434719	AF434719 Buchnera

ALIGNMENTS

RESULT 1
LIU45241
LOCUS

DEFINITION

LIU45241

Lawsonia intracellularis plasmid pISI-2

gene, partial cds, L27 50S ribosomal binding protein homolog, L21

50S ribosomal binding protein homolog, GroES/HSP10 homolog, and

GroEL/HSP60 homolog genes, complete cds.

U45241

U45241.1 GI:35999919

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 4710)

4710 bp DNA linear BCT 23-SEP-1998
Lawsonia intracellularis plasmid pISI-2 GTP-binding protein homolog
gene, partial cds, L27 50S ribosomal binding protein homolog, L21
50S ribosomal binding protein homolog, GroES/HSP10 homolog, and
GroEL/HSP60 homolog genes, complete cds.

Lawsonia intracellularis
Lawsonia intracellularis
Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacterales;
Desulfobacterales; Lawsonia.

Date, C.J., Moses, E.K., Ong, C.C., Morrow, C.J., Reed, M.B., Hasse, D. and Strugnell, R.A.
 Identification and sequencing of the groE operon and flanking genes of *Lawsonia intracellularis*: use in phylogeny
 Microbiology 144 (Pt 8), 2073-2084. (1998)
 98386497
 9720028
 2 (bases 1 to 4710)
 Dale, C.J.H.
 Direct Submission
 Submitted (11-JAN-1996) C. Jane H. Dale, Molecular Biology, Victorian Institute of Animal Science, 475 Mickleham Road, Attwood, VIC 3049, Australia
 Location/Qualifiers
 1..4710
 /organism="Lawsonia intracellularis"
 /mol_type="genomic DNA"
 /db_xref="taxon:29546"
 /plasmid="pSi-2"
 complement(1..345)
 /codon_start=1
 /transl_table=1
 /product="GTP-binding protein homolog"
 /protein_id="AAC36496.1"
 /db_xref="GI:3599920"
 /translation="MRFVDEVTISVSAGKGGNGCVSPRRKFIKGGPENGDDGGGNIFKADSRLLLYDFRVQRHYRANQNGEGGKSGRHKKGEDLILHLPVGTIIPQLLDKEYFLVLDLRPGV"
 complement(508..774)
 /codon_start=1
 /transl_table=1
 /product="L27 50S ribosomal binding protein homolog"
 /protein_id="AAC36497.1"
 /db_xref="GI:3599921"
 /translation="MAHKAGGSSRNGRSDPGRRGIKRGQXKVLGNILVRLGTLKIHPENGWGRDYTLFAKNGIVYTYFRKCKYSKVHIQPTV"
 complement(787..1095)
 /codon_start=1
 /transl_table=1
 /product="L21 50S ribosomal binding protein homolog"
 /protein_id="AAC36498.1"
 /db_xref="GI:3599922"
 /translation="MNAIKPLNDRVLVRLSEKTAGGLYIPDTAKBKPRGEVAVGPGKHTDDKGLIPMAVKAGDTVLFNKYAGTEVKLDGVBEHLVMREDDIILAVITGTRK"
 2048..3694
 /note="heat shock protein; chaperonin"
 /codon_start=1
 /transl_table=1
 /product="GroEL/HSP60 homolog"
 /protein_id="AAC36500.1"
 /db_xref="GI:3599924"
 /translation="WASKELIFDFAKAREKL SRGVDDKLANAVKVTLPKGRNVVIEKSF GSPVTKDGVSAKEIELEBDFENMGAMQVAPKTSDIAGDGTATVLAQAIYRE GVKLVAGRNPMALRGKIDKAVAYTKELSDIKPDRDQKEIAQVGTISANSDTIGN IIAEMAKVGGVITVEBAKGLLETTLDVVEGKFDRIYSPYFVTPNPKVVCLELNP YILCNKKITSGKMDPLILQVAKNRPPLIIIAEDVEGEALATLVNKLKRALQVAV KAPGFERRKAMLEDIATLTCGEAIFEDRGIKLENVSSLSIGTAKRVVIDKENTIVD CAGKSEDIKARKVQIRAOIEBTSYDREKLQERLAKLVGVAVIHVGAATEEMKEK KDRVEDALNATRAVEGIVPGGTA FVRSIKVLDDIKPADDDLAGLNIIRLSLEP LKQIAAALVEGSIYVKEVREPKDPGRFNAAAGEYEDLIKAGVIDPKVTRIALQNA SVASLLITTECAIAEKEPDKMDPMPGGGGMGGMDGMV"
 3821..4702
 /note="orf; similar to *Escherichia coli* possible


```
RESULT 4
LOCUS LPNHTPBG
DEFINITION Legionella pneumophila htp heat shock operon HtpB (hptB gene), complete cds.
ACCESSION M31918
VERSION M31918.1 GI:149691
KEYWORDS htp heat shock operon.
SOURCE Legionella pneumophila
ORGANISM Legionella pneumophila
REFERENCE 1
AUTHORS Hoffman, P.S., Houston, L. and Butler, C.A.
TITLE Legionella pneumophila htpB heat shock operon: nucleotide sequence and expression of the 60-kilodalton antigen in L. pneumophila-infected HeLa cells
JOURNAL Infect. Immun. 58 (10), 3380-3387 (1990)
MEDLINE 90382960
PUBMED 2205580
COMMENT Original source text: Legionella pneumophila (strain SVir) (library: cosmid clone pH79) DNA.
Draft entry and computer-readable sequence for [Unpublished (1990) Univ. of TN at Memphis, Memphis TN 38163] kindly submitted by P.S. Hoffman, 05-FEB-1990.
FEATURES
     source
         1..1811
             /organism="Legionella pneumophila"
             /mol_type="genomic DNA"
             /strain="SVir"
             /db_xref="taxon:446"
             /tissue_lib="cosmid clone pH79"
             /gene="htpB"
             /gene="htpB"
             /codon_start=1
             /transl_table=11
             /protein_id="AAA25299.1"
             /db_xref="GI:149692"
             /translation="MAKELRFDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYG
             APTVTKDVSVAKEIEFEHFRFMNGAMVKEVASKTSIDTAGDGTATATVLAISILVEG
             HAAAGNPMDLKRGIDKVLATKLOAMSKPKDSKAIQVCTIISANSDEALGAI
             IIAEMEKVYKQGVLTVEDNGLENELYVEGMDPQGYISPYFINNQNMSCLEHPF
             ILLVDKVVSIKREMLSVLEGVAKSRPILLIADVEGALATILVNNMKGIVKVCVK
             APGDRRKLMDQIALLTKGVISEEIGKSLLEGATLEDLGSARKIVVTKENTTLLIDG
             EGKATEINARIAQIRAEQMEETSDYDREKLQERVAKLAGVAVIKVGAATEVEMKEK
             ARVEDALHATAAEVGGIVAGGVVALIRAKALDSLKGDNDQNMGINILRAIRBPM
             RQIVTNAGYEASVVVNVKVAHKDNYGFNAATGEYGDVEMGILDPTKVTMALQNAAS
             VASLMLTTECMVADLPKKEEGVAGDMGMGMGMGMGMGM"
             repeat_unit 1699..1780
             /rpt_type=inverted
BASE COUNT 531 a 299 c 448 g 533 t
ORIGIN
Query Match 47.3%; Score 778.4; DB 1; Length 1811;
Best Local Similarity 67.7%; Pred. No. 1.1e-128;
Matches 1107; Conservative 0; Mismatches 521; Indels 6; Gaps 1;
QY 8 CTAAGAAATCCTTTTGTATGCTAAAGCCCGGTGAAACATTTTCACGAGGTGTAGATAAAC 67
DB 38 CTAAGAAATACGTTTGTGTATGCTACGCTCGCTACAAATGCTTCTGCTGTATATGCAT 97
QY 68 TTGCAATGCTCTTAAGTAACACTTGACCTAAAGCCCGTAATGCTGTTATTTGAAAGT 127
DB 98 TAGCAGATGCGGTTCAGTTACTATGCGTCCACGTCGCGCGTAATGTTTATTTGAAAAAT 157
QY 128 CTTTGGTTCCTCCAGTTATTACAAAGATGGTGTATCTGTGTCGAAAGAAATTTGAAGT 187
DB 158 CTTATGGCGCTCTACTACTACTAAAGACGTTGTCTGTGTCGCAAGAAATTTGAATTG 217
QY 188 AAGATAAGTTTGAAATATGGCGCTCAATGGTTTAAAGAGTAGTCCCAAACTAGCG 247
DB 218 AGCATCGTTTTCATGAACATGGCGCTCAATGGTTTAAAGAGTAGTCCCAAACTAGCG 277
QY 248 ATATTGCTGCTGATGAACTACAAACAGCAAGTCTCTGCACAGCTATTTATTCGTGAAG 307
DB 278 ATATGCTGCTGATGTTACTACTACTGCAACAGTATTTGCTGCTCTTATTTCTTTGTAAG 337
QY 308 GTGTAAACCTTTAGCAGCTGGTCTGTAATCTTATGCCATTTAAACGTGGCATAGATAAG 367
DB 338 GTCAACAAAGCAGTTGCTGCTGTATGAATCCAATGGATCTCAACCGCGTATTGATAAAG 397
QY 368 CTGTGTTGCTCTTACTAAAGAACTAAGCAGATTAACAAGCCTACTCTGTCACCAAAAG 427
DB 398 CAGTATTAGCAGTTACCAAAAATTTACAAGCTATGCTTAAGCCATGCAAGACAGCAAG 457
QY 428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGTATACAAACATPAGTAATATCATAG 487
DB 458 CCAATTGCTCAAGTTGGAACCATTTCTGCTTAATCCGATGAAGCATTTGGTGTATCATTTG 517
QY 488 CTGAAGCTATGCTAAAGTTTGAAGAGGAGTGTATACAGTTCAGGAGCTAAAGGTC 547
DB 518 CTGAAGCAATGGAAGAGTTGTAAGAGGTTTATTAACCGTTGAAGATGTTAATGAT 577
QY 548 TTGAAACTACATTAGATGTTGTAAGAGTAATGAAGTTTGCACCGTGGCTACCTCTCTCCAT 607
DB 578 TCGAAAATGAGCTTTATGTTGTTGAAGGTATGCAATTTGATCCGCTTACATTTCTCCAT 637
QY 608 ACTTTGTAACATACTCTGAGAAAATGGTTTGTGAACCTTGTATTAACCCCTTATATCCCTTTGTA 667
DB 638 ACTTCATCAACCAACGAGCAAAACATGAGCTGTGAACCTTGAGCATCCATTCATTTATTGG 697
QY 668 ATGAGAAAAGATTACTAGCATGAAGACATCTCTACCAATCTTAGAACAAGTTGCTTAAG 727
DB 698 TTGCAAAAAGATTTTCCAGTATTCGTTGAAATTTGTCGGTATTTGGAAGGTTTGTCTAAAT 757
QY 728 TAAACCGTCCACTCTCTTATTTATGCTGAAGAGCTAGAAGTCAAGCAGTTCGCAACACTTG 787
DB 758 CTGGCGCTCTTGTGTTGATCATCGCAGAGATGTTGAAGCGAGAGCTTTAGCTACTCTGG 817
QY 788 TAGTCAATAAGCTCCGTGAGCAGCATCCCAAGTTGTAGCCGTGTAAGAGCTCTCTGTTTGGTG 847
DB 818 TAGTCAACAACATGCGCGGTATTTGTAAGAATATGCTGTCAAGCGCTGCTGTTTGGTG 877
QY 848 AACCGCTTAAGCTATGCTTGAAGATATGCTATCTTACTCTGAGAGAGCAATATTTG 907
DB 878 ATCTGCGCAAGCGATGTTTGAAGACATTTGCTATTTTGAAGGCTCAAGTTATTTCTG 937
QY 908 AAGATCGTGGTATAAGCTTGAAGATGTAAGCTTCTCTTCTTTAGCAACAGCTAAACGTTG 967
DB 938 AAGAAATTTGCAAGAGCTTGAAGAGTCTACTCTGGAAGATCTTGTAGTCTTAAGCGNA 997
QY 968 TAGTTATTGACAAAGAAATACTACTATGCTGTGTTGTTGGAAGATTCAGAAAGATATTA 1027
DB 998 TCGTTGTTTACCAAGAAACACTACTATCAITGATGTTGTAAGAAAGGCACTGAAATTA 1057
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAACAGCTCAGATATGATCGTG 1087
DB 1058 ATGCTCGATTGCTCAAAATTCGTGCACAAATTTGAAGAAACCACTTCTGATTACGATAGAG 1117
QY 1088 AAAAATTTCAAGAACGCTCTTCAAAACCTTGTGTTGAGTAGTCTTATCCATGTTGGAG 1147
DB 1118 AAAAATTTCAAGAGCGCTTCTAAGTTAGCTGGTGGTGTGCTGTTATCAAGATTGGCG 1177
QY 1148 CTGCTACTGAACTGAAATGAAGAGAGAGGATCGTGTAGAGATGCTCTTAATGCA 1207
DB 1178 CTGCCACAGAAAGTTGAAATGAAGAGAGAGAGAGTCTGTTGAGATGCTCTTTCATGCTA 1237
QY 1208 CAAGAGCTGCGGTTGAAAGAGGATTTGTCCTCTGTTGGTGGTACTCTTTTTCGCTCCA 1267
DB 1238 CTCGCGCTGAGTAGAAGAGGATTCGTTGCGGTTGGGCTTGCCCTTGATTCGTCGTC 1297
QY 1268 TTAAGTCTCTTGTATGATATTAACCTGCTGATGATGATGAATGAACTTGTGCTTAATATCA 1327
```

```

1298 AGAAAGCTCTGATTCATTAAGAGTGATTAATGACGATCAAAATATGGGTATCAATATTT 1357
1328 TCCGTCGTTCTCTGAAGAGCCCTTACGTCACAAATGCTGCAATGCTGCTATGAAGTT 1387
1358 TAGGTCGCGCTATTGAATCTCAATCGCTGAGATTTGTTACTAAGCAGGATATGAAGCTT 1417
1388 CTATTGTTGTAGAAAAGTTCTGTAACCAAAAGATGGTTTGGATTTAATGCTGCATCAG 1447
1418 CTGTTGTAGTAACCAAGTAGCTGAGCACAAAGACAACTACCGTTTCAACGCTGCAACTG 1477
1448 GAGATATGACGACCTTATTAAGCTGGTGTCAATTCATCTCAAAAGATTACACGCTATTG 1507
1478 GTGATACGCTGATATGTTGAGATGGGTATCTTGAATCAACTAAAGTAAACCGGTATGG 1537
1508 CATTACAAATGAGCATCAGTAGCTCTCTTACTTCACTACAGAAATGCGGTATTGCTG 1567
1538 CTCTGCAAAATGAGCTTCTGTAGCCAGTTTGATGTTGACTACTGAATGATGGTTGCTG 1597
1568 AAAACACGAACTAAAAGATATGCTATGCTGCGGTGATGGTGGTATGGTATGGTGC 1627
1598 A-----TCTGCTTAAAGAAAGAGAGGTGTTGGTCCCGCGATATGGCGCGCATGGCG 1651
1628 GATGACGCTATG 1641
1652 GAATGGGTGCATG 1665

```

```

RESULT 5
AB057417
LOCUS AB057417 2715 bp DNA linear BCT 16-MAR-2001
DEFINITION Pseudoalteromonas sp. PSIM3 gene for GroES, GroEL, complete cds.
ACCESSION AB057417
VERSION AB057417.1 GI:13366171
KEYWORDS
SOURCE Pseudoalteromonas sp. PSIM3
ORGANISM Pseudoalteromonas sp. PSIM3
REFERENCE 1 (sites)
AUTHORS Kurusu,Y. and Nakamura,T.
TITLE Molecular chaperone of Psychrophile
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2715)
AUTHORS Kurusu,Y.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2001) Yasuou Kurusu, Ibaraki University School
of Agriculture, Laboratory of Molecular Microbiology; chuo 3-21-1,
Inashiki Ami, Ibaraki 300-0393, Japan
(E-mail:krsy@ipc.ibaraki.ac.jp, Tel:81-298-88-8646,
Fax:81-298-88-8646)
FEATURES
Location/Qualifiers
1..2715
/organism="Pseudoalteromonas sp. PSIM3"
/mol_type="genomic DNA"
/strain="PSIM3"
/db_xref="taxon:87791"
/country="Japan"
/notes="isolated from the Japan Trench"
376..388
383..388
402..410
406..411
458..761
/gene="groES"
458..464
/gene="groES"
474..761
/gene="groES"
/codon_start=1
/transl_table=1
/product="GroES"
/protein_id="BAB39464.1"

```

```

promoter -35_signal
promoter -10_signal
gene
RBS
CDS

```

```

/db_xref="GI:13366172"
/translation="MNIRPQDRVIVKRLLEETKSGAGIVLTCGSAEAKSTRGVAVG
NGRIIDNGDVRALEVKGAGTDLFGSVTEKTEIQEYILIMREDNLTGIVG"
797..2460
/gene="groEL"
RBS 797..803
/gene="groEL"
811..2460
/gene="groEL"
/function="molecular chaperone"
/codon_start=1
/transl_table=1
/product="GroEL"
/protein_id="BAB39465.1"
/db_xref="GI:13366173"
/translation="MAAKEVLEFAGDARAKMLTGVNLANAQVTLGPKGRNVVLKDSF
GSPVITKGVSAKEIETLEDFENMGAMQVKEVASKANDAAAGDGTATATVLAQSIYNE
GLKVAAGNPMDLKRGIDKAIIAVAELKALSPCADTKIAQVGTISANSKEIJD
IIAEMEKVGRNTGVTIVBEGOSLENELDWEGMDFRGYSPLFINNAEKGAVELDN
PTLLVDKVSINIRELLPTLEAVAKASPLIIAEDLEGEALATVNNMRGIVKVS
VKAPGDRKAMLODIAVLGTGTVISBEIGLELEKATVEDLGATKRVVITKDDTII
DGAGEEGITGRVSIKQIIEBATSDDYDKELQERMAKLAGGVAVIKVGAATEIEMKE
KKDRVEDLUHTRAAVEGVVPGGVALVRAASKVELLEDNEDQNHGVKVALAMEA
PLQIVTNAGDEASVVINAVKGGEGNYNAATGEYNDMEIMGILDPTKTVRSALQFA
GSTAGLMTITEAMVAEIPKEEAAGAPDMGGMGMGMGMGM"
terminator 2519..2567
BASE COUNT 860 a 513 c 613 g 729 t
ORIGIN
Query Match 47.2%; Score 777; DB 1; Length 2715;
Best Local Similarity 67.4%; Pred. No. 1.8e-128;
Matches 1111; Conservative 0; Mismatches 535; Indels 3; Gaps 1;
QY 1 ATGCGTCTTAAGAAATCCTTTTGTATGCTAAAGCCGCGTAAAGCTTTTCAAGAGGTGA 60
DB 811 ATGCGAGCTAAAGAGTACTTTTTCAGGTCGACGCGCTAAATGCTAACTGGCGTA 870
QY 61 GATAAACTTGCAAATGCTTTAAAGTAAACACTTGGACCTTAAAGCCGCTAAATGCTGTTAT 120
DB 871 ACATCTAGCAACCGCAGTTAAAGTTACATTAAGTCTTAAGCCGCTAACGTTGACTA 930
QY 121 GAAAAGTCTTTTGGTTCCTCCAGTTTATACAAAAGATGCTGTTATCTGTTGCAAAAGAAAT 180
DB 931 GACAAATCGTTTGGCTCTCCAGTGATCACTAAAGATGCTGTTATCTGTTAGCAAAAGAAAT 990
QY 181 GAACCTGAAGATAGTTTGAATATGGCGCTCAATGCTTAAAGCTTAAAGAGTAGCTCCCAA 240
DB 991 GAACCTGAAGATAGTTTGAATATGGCGCTCAATGCTTAAAGAGTAGCTTAAAGAGTAGCT 1050
QY 241 ACTAGCGATATTGCTGGTATGGAACCTACAAACAGCAACAGTCCCTTGCACAGCTTATTAT 300
DB 1051 GCAATGATCGCGCGCGATGTTACAACTACCGCTACAGTCTTGCACAGCTTATTATTT 1110
QY 301 CGTGAAGCTTAAACCTTTAGCAGCTGGTGGTAACTCTATGCGCATTAACAGTGGCAT 360
DB 1111 AATGAAGCTTAAAGCGTTTGTGCGAGCATGAACCAATGAGACCTTAAACGTTGTTAT 1170
QY 361 GATTAAGCTTGTGTTGCTGTTTACTAAAGAACTAAGCGACATTAAGAGCTTACTCTGTCAC 420
DB 1171 GATAAGCAATCATTTGCTGTTTGCAGAGCTTAAAGCGCTTCTGTTCTTGTGCTGAT 1230
QY 421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAAACAATAGGTAAT 480
DB 1231 ACTTAAGCTATTGCTCAAGTAGGTACTATTTCAGTAACTCTGATAAGAAATTTGGCAG 1290
QY 481 ATCATAGCTGAAGCTATGGCTTAAAGTTG---GAAAAGAGGTGTTATCACAGTTGAGAA 537
DB 1291 ATCATTCAGAGCAATGGAAGAAAGTAGGCGTAAATACAGGTGTTATTTACTGTTGAGAA 1350
QY 538 GCTAAAGCTCTTGAACATACATTAGATGTTGTTGAAGGAATGAAGTTGACCGTGGCTAC 597
DB 1351 GGTCAATCACTAGAGAACGAACTAGATGTTGTAGAGGCGATGCGAGTTTGACCGTGGTTAC 1410

```


[illegible]

11301 Wilshire Blvd., Los Angeles, CA 90073, USA

FEATURES

Location/Qualifiers

1..1644

/organism="Bartonella henselae"

/mol_type="genomic DNA"

/strain="ATCC 49882"

/db_xref="taxon:38323"

1..1644

/gene="groEL"

1..1644

/gene="groEL"

/note="GroEL"

/codon_start=1

/translation=1

/product="heat shock protein HSP60"

/protein_id="AAB69094.1"

/db_xref="GI:2358234"

/translation="MAAEKVFQREARELLRGLDILANAVKVLTLGKGNVVIDKSF
GAPRIKDVSVAKETELDKPENMGAOMLREVASKTNDIAGDGTATATVLAIVQVE
GVKAVAGMNPMLKRGIDANDEVVANLFKKAQKIQTSAEIAQVCTI SANGAAEIGK
MIADAMEKVNNEGVIITVEAEIETEVVEGQFDRGLSPFYVNAEKVADLDDP
YLIHKKLSNLSQSLPVLAEVQSGKPLIIIAEDVEGALATLVNKLKRGKLIKIAV
KAPGDFDRKAMLEDTAILTSQVISEDVIGKLENTVLDMLGRKKNIS KENTIID
GAGQSEINARVNOIKVQIEETSDYDREKLOERLAKLAGVAIVRGATVEVEKEK
KORVDALNATRAAEVIGVAGGTALLRAANALTVKGSNPDOEAGINIVRRALQAPA
RQIATNAGEBAALIVKVLNENADTGYNTATGEFGDLIALGIVDPVKVRSALQNA
SIASLIITTEAMVAEVPKDDTPVPPMPFGSGMGGMGDF"

BASE COUNT 486 a 309 c 419 g 430 t

ORIGIN

Query Match 46.8%; Score 770.6; DB 1; Length 1644;

Best Local Similarity 68.1%; Pred. No. 2.7e-127; Indels 9; Gaps 3;

Matches 1118; Conservative 0; Mismatches 514;

QY 1 ATGGCTCTTAAAGAAATCCCTTTTGTATGCTTAAAGCCCGTGAAGAAACCTTCCAGAGGTGA 60
DB 1 ATGGCTCTTAAAGAGTGAAGTTTGGCGGTGAAGCGGTGAGCGTTTGTGGCGGGGTT 60
QY 61 GATAAATCTGCAATGCTGTTAAAGTAACTTGGACCTAAAGCCGTAATGCTGTTATT 120
DB 61 GATATCTTCTGTAACGCTGTTAAAGTAACTTGGACCTAAAGCCGTAATGCTGTTATT 120
QY 121 GAAAGTCTTTGGTCCCGAGTATTACAAAGATGCTGATCTGTTGCAAGAAAT 180
DB 121 GATAAATCTTGGTGGCGCTCGCAATACAAAGATGCTGATCTGTTGCAAGAAAT 180
QY 181 GAACCTTGAAGATAAGTTGAAAATATGGCGCTCAAAATGTTTAAAGAGTAGCTCCCAA 240
DB 181 GAACCTTGAAGATAAGTTGAAAATATGGTGGCGCAATGTTGCGTGAAGTTGCTTCTAA 240
QY 241 ACTAGCGATATGCTGTGATGGAACCTAACACAGCAACAGTCCTTGCACAGCTATTAT 300
DB 241 ACCAATGACATGCTGTGATGGAACCAACACAGCAACAGTCCTTGTAGACAGGCTATT 300
QY 301 CGTGAAGGTGAAACTTTGAGAGCTGCTGCTAATCTATGGCCATTAAGAGTGGCATA 360
DB 301 CAAGAAGGTGAAAGCGGTGCTGCGAGGCATGACCAATGATCTCAACAGTGGTATC 360
QY 361 GATAAAGCTGTTGTTGCTTACTTAAAGAACTAAGCGACATTAACAAAGCTTACTCGT 420
DB 361 GATGCTGTTGATGAGTGGTGGCAATCTTTTCAAAAGCGAAGAAATCCAAACT 420
QY 421 CAAAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAAATAGGTA 480
DB 421 TCAGCAAGAAATGCAAGTGGGCAAAATTTCCGCTAATGCTGCTGCTGAAATCGGCAA 480
QY 481 ATCATAGCTAGCTATGCTGCTAAAGTTGGAAGAGGTGTTATCACAGTTGAGAGCT 540
DB 481 ATGATCGCTGATGCTGCTGGAAGAAATTTGGCAATGAAGGTGCTATACCGTTGGAAGA 540
QY 541 AAAGGTCTTGAACATACATAGATGCTGTTGAAGGAATGAAGTTGACCGTGGCTACCT 600
DB 541 AAAACTGCTGAACGGAATGAGAGTCTTGAAGGAATGAGTTGATCGTGGATATCTT 600


```
RESULT 9
PHA243594          2241 bp  DNA  linear  BCT 05-FEB-2001
LOCUS              Pseudoalteromonas haloplanktis groEL, groES genes.
DEFINITION         AJ243594
ACCESSION          AJ243594
VERSION            GI:12697197
KEYWORDS            groEL gene; groES gene; GROES protein.
SOURCE             Pseudoalteromonas haloplanktis
ORGANISM            Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
                    Alteromonadaceae; Pseudoalteromonas.
REFERENCE
AUTHORS            Tosco, A., Birolo, L., Scaloni, A., Sannia, G. and Marino, G.
TITLE              A GroEL-like protein from the psychrotrophic bacterium
                    Pseudoalteromonas haloplanktis TAC125
JOURNAL            Unpublished
REFERENCE
AUTHORS            Tosco, A.
TITLE              Direct Submission
JOURNAL            Submitted (12-JUL-1999) Tosco A., Organic and Biological Chemistry,
                    University of Naples, via Mezzocannone 16 (NA), I-80134, ITALY
REMARK             Revised by [3]
                    3 (bases 1 to 2241)
AUTHORS            Stefania, M.
TITLE              Direct Submission
JOURNAL            Submitted (05-FEB-2001) Stefania M., Organic and Biological
                    Chemistry, University of Naples, via Mezzocannone 16 (NA), I-80134,
                    ITALY
COMMENT            On Feb 6, 2001 this sequence version replaced gi:5524757.
FEATURES
    source          1..2241
                    /organism="Pseudoalteromonas haloplanktis"
                    /mol_type="genomic DNA"
                    /strain="TAC125"
                    /db_xref="taxon:228"
    gene            71..133
                    /gene="groESL"
    -35_signal      71..83
                    /gene="groESL"
                    /evidence=experimental
    -10_signal      97..105
                    /gene="groESL"
                    /note="sigma 32-dependent promoter"
                    /evidence=experimental
    -35_signal      105..110
                    /gene="groESL"
                    /evidence=experimental
    -10_signal      128..133
                    /gene="groESL"
                    /note="sigma 70-dependent promoter"
                    /evidence=experimental
    gene            153..455
                    /gene="groES"
    RBS             153..455
                    /gene="groES"
    CDS             168..455
                    /function="molecular co-chaperone"
                    /codon_start=1
                    /transl_table=11
                    /product="groES protein"
                    /protein_id="CAC28359.1"
                    /db_xref="GI:12697198"
                    /translation="MNIRPLHDRVIVKLEETKSGIGVLTGSAEKSIRGEVVAVG
                    NGRILESDVPALEVKAGDVTLFSGYVEKVEKIEGQYILIMREDNILGIVG"
    gene            489..2147
                    /gene="groEL"
    RBS             489..493
                    /gene="groEL"
    CDS             504..2147
                    /gene="groEL"
                    /function="molecular chaperone"
```

```

/codon_start=1
/transl_table=11
/product="GroEL protein"
/protein_id="CAB50775.1"
/db_xref="GI:5524758"
/db_xref="SPTREMBL:Q9XAU7"
/translation="MAAKEVLFAGDARAKMLTGVNITLANAVKTLGPKGRNVVLKSF
GSPVITKDGVSVAKEIELEDFENMGCAOMVEKVASKANDAGDGTATTATVLAGSIYNE
GLKAVAGMNPMDLRGIDKAVIAVAELKALSPVCSDDTKAIAQVGTIISANSKEIGD
IIAQAMEKVRNSGVITVEEGOSLENELDVGMQPDRLGYLSPYFINSPEKGTVELDN
PILIVDKKISINRILLPTLEAVAKSKPLLIIAEDLEGEALATLVNMRGIVKVSVA
VAPGFGRRKAMQIDIAVITGGTITSEIEGLELEKATVEDLGTARVITTKDDTII
DGAGEEAGIRGVSOIKAOIEATSDYDEKLEQERMAKLAVGAVIKVGAATEMKE
KKDRVEDNALTRAAVEGVPGGVALVRRAKLVLDVGDNDONHGKVALRAMEA
PLRQIVTNAGDEASVINAVKAGSNFGYNATGEYNDMEIGILDPTKVTRSALQFA
GSIAGLIMITTEAMVAIEIPKDDSPDMGGMGGMGGM"
2178..2202
/gene="groESL"
2178..2202
/gene="groESL"
terminator
BASE COUNT 735 a 416 c 511 g 579 t
ORIGIN
Query Match 46.4%; Score 764.6; DB 1; Length 2241;
Best Local Similarity 67.4%; Pred. No. 3e-126;
Matches 1111; Conservative 0; Mismatches 529; Indels 9; Gaps 2;
QY 1 ATGGCTTAAAGAAATCCTTTTGTATGCTAAAGCCGTGAAAACTTTTCAGAGGTGA 60
DB 504 ATGGCAGCAAAAGAGTACTTTTTCAGGTGACGCGCGCTAAATGCTAACTGCGTA 563
QY 61 GATAAATCTGCAATGCTGTAAAGTAACTAACACTTGACCTTAAAGCGCGTAATCTTAT 120
DB 564 AACATCTTGGCAACGAGTAAAGTTACATTAGTCTTAAAGCGCGTAACTGTTACTA 623
QY 121 GAAAAAGTCTTTGGTTCCTCCAGTTATTACAAAGATGGTGTATCTGTGCAAAAGAAAT 180
DB 624 GATAAATCATTTGGCTCTCCAGTTATCACTAAAGATGGTGTATCTGTAGCAAAAGAAAT 683
QY 181 GAACCTTGAGATAGTTTGAATAATATGGCGCTCAATATGGTTAAAGAGTAGCTCCAAA 240
DB 684 GAACCTTGAGATAGTTTGAATAATATGGCGCTCAATATGGTTAAAGAGTAGCTCCAAA 743
QY 241 ACTACGATATTTGCTGGTGTAGTGAATCAACAGCAAGTCTCTTGGCAAGCTATTTAT 300
DB 744 GCTAACGATGACGCGCGGATGTTACAACTACCGCTACAGTACTTGCACAGTCTATTGA 803
QY 301 CGTGAAGGTGAAAACTTTGTAGCAGCTGGTCTGTAATCTCTATGGCCATTAAGCGTGA 360
DB 804 AACGAAGGCTTAAAGCTTTGACAGCGCATGAACCCAAATGGACCTTAAAGCGCGCATC 863
QY 361 GATAAAGCTGTGTGCTGTACTTAAGAACTAAGCACTAAGCGCATTAAGAGCTACTCGTGA 420
DB 864 GACAAAGCAGTAATTTGCTGTGTTGACAGCTTAAAGCGCTTATCTGTCTCCATGCTGAC 923
QY 421 CAAAAAGAAATAGCTCAAGTTTGAACCACTTCTGCAAACTCTGTATACAACTAGGTAT 480
DB 924 ACTAAGCAATGACAACTAGTACTTATTTACGTTAACTCTGTATTAAGAGATTGGCAG 983
QY 481 ATCATAGCTGAAGCTATGGCTTAAAGTTG ---GAAAGGAGGTGTTTATCACAGTTGAGAA 537
DB 984 ATCATGCAAGCAATGGAAGAAAGTAGTGGCGTAACTAGCTGGTGTATTACTGTAGAGAA 1043
QY 538 GCTAAGAGCTTGAAGAACTACATTAGATGTTGGTTGAGAGAAAGAGTTTGACCGTGGCTAC 597
DB 1044 GGTCAATCAGTAGAATAATGAACTGATGTTGTTGAAGGCGATGCGATTTGACCGTGTAC 1103
QY 598 CTCTCTCCATCTTTGTAACCTAACTTCTGAGAAATGGTTTGTGAACTTGTAAACCTTAT 657
DB 1104 CTATCTCTTACTTTTATTAACAGCCCAAGAAAGGTACTGTTGAACTAGATTAACCAATTT 1163
QY 658 ATCCTTTGTATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAA 717
|||||
```



```
Db 1164 ATTCTACTTGTAGATAAAAAAATATCTAATTCGCGAATTATTACCAATAGAGCA 1223
Qy 718 GTTGCTAAAGTAACCGTCCACTCTTATTTATGCTGAAGACGTAGAGGTGAAGCACTT 777
Db 1224 GTTGCTAAAGCAAGCAAGCCACTACTAATCATCGCAAGACCTTGAAGCGCAATTA 1283
Qy 778 GCAACACTTGTAGTCAATAGCTCCGCTGGGACACTCCAGTGTGTAGCCGTAAAGCTCT 837
Db 1284 GCTACATTTAGTATTAATTAACATGCTGGTGTATTTGTTAAAGTATCAGCATGAAGCTCT 1343
Qy 838 GGTGTTGGTGAACCGCTGAAAGTATGCTTTGAAGATATTGCTATCTTACTGAGAGAA 897
Db 1344 GGTGTTGGCGACGCTGTAAGCAATGCTACAGATATCGCTGTTAATGTCGCGTACT 1403
Qy 898 GCAATATTTGAAGATCGTGGTATAAAGCTTTGAAATGTAAGCTGTCTTTTAGAACA 957
Db 1404 GTGATTTGAGAAATCGGCTTGTAGCTTGAAGAGCAACAGTTGAAGACCTTAGGTACA 1463
Qy 958 GCTAAAGCTGTAGTTTACAAAGAAATACACTACTCTGTTGATGCTGCTGGAATCA 1017
Db 1464 GCTGAGCGGCTAATTAATCACTAAGATGATACAACTAATCATTTGCTGCTGCGAAGAA 1523
Qy 1018 GAAGATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAAGAAACAGCTCAGAT 1077
Db 1524 GCAGGTATTAACGCGCTGTTTCAAAATTAAGACACAAATCGAAGCAACATCAGAC 1583
Qy 1078 TATGATCGTGAAGAACTTCAAGAACGCTTTGCAAACTTTGTTGGTGAGTAGCTTTATC 1137
Db 1584 TACGATAAAGAAAACTTACAGAGCGCATGGCTAAACTAGCTGCGGCTGTTGAGTAATC 1643
Qy 1138 CATGTTGGAGCTGCTACTGAACTGAATCAAGAGAGAGAGATCGTGTAGAGATGCT 1197
Db 1644 AAGTGTGGTCTCGACTGAATGGAATGGAATGGAAGAGAAAGAACCGCTGTTGAGATGCA 1703
Qy 1198 CTAATGCAACAGAGCTGCGGTTGAAGAGAGTATTGCTCCCTGGTGTGCTGCTGTTTT 1257
Db 1704 TTAATGCAACTGTCGAGCGGTTGAAGAGGTGTAGTACTTGGCGCGCGTTGCACTA 1763
Qy 1258 GTCGCTCCATTAAGTCTTGTATGATTAATTAACCTGCTGATGATGATGAATGCTGGA 1317
Db 1764 GTTCGTGACCAAGTAAGCTGGTTGATTTAGTTGGGACAAACGAAGATCAAAACCAAGT 1823
Qy 1318 CTTAATATCATCGTCTGTTCTTGAAGACCTTTAGTCAAAATTCGTGAAGTCTGCGC 1377
Db 1824 ATTAAGTTCACACTTCGTGCAATGGAAGCGCCACTTCGTCAAAATCGTAACAAACGCAAGT 1883
Qy 1378 TATGAAGTTCTATTGTTGTAGAAAAGTTTCGTGAACCAAAAGATGTTTGGATTATTAAT 1437
Db 1884 GACGAGCTTCAGTTGTTAATTAACGCAAGTTAAGCAGGTTTCAGTAATTTGTTACAAT 1943
Qy 1438 GCTGCTCAGGAGATTAAGACCTTATTAAAGCTGGTGTATGATGATCTTAAAGAAATTT 1497
Db 1944 GCTGCTACTGGCAATCAATGATGATTAATGAAATGGGTATCTTAGATCAACCTAAAGTA 2003
Qy 1498 ACAGTATTCATTAACAAATGACATCAGTACGCTCCTTACTTCTAATCTACAGATGTC 1557
Db 2004 ACAGCTCTCAGCTACAAATTTGCGGGTCTATTGCGGGTCTTAATGATCATTACCGAAGCT 2063
Qy 1558 GCTATTGCTGAAAAACCAAGAACCTTAAAGAGATATGCTATGCTGCGGTGGTATGGGT 1617
Db 2064 ATGTAGCTGAAA-----TCCCTTAAGATGATAGCGCTCCTGATATGGGTGGCATGGC 2117
Qy 1618 GGTATGGGTGATGAGCGGTATGACTA 1646
Db 2118 GGTATGGGTGAATGGCGGTATGATGTA 2146
```

RESULT 10
RQU78515
LOCUS Bartonella quintana heat shock protein HSP60 (groEL) gene, complete cds.
DEFINITION
ACCESSION U78515

```
U78515.1 GI:2358235
VERSION 1
KEYWORDS Bartonella quintana
SOURCE Bartonella quintana
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bartonellaceae; Bartonella.
REFERENCE 1 (bases 1 to 1644)
AUTHORS Haake, D.A., Summers, T.A., McCoy, A.M. and Schwartzman, W.
TITLE Heat shock response and groEL sequence of Bartonella henselae and
Bartonella quintana
JOURNAL Microbiology 143 (Pt 8), 2807-2815 (1997)
MEDLINE 97419519
PUBMED 9274034
REFERENCE 2 (bases 1 to 1644)
AUTHORS Haake, D.A., Summers, T.A. and Schwartzman, W.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-1996) Medicine, 111F, West Los Angeles VAMC,
11301 Wilshire Blvd., Los Angeles, CA 90073, USA
FEATURES
source
1..1644
/organism="Bartonella quintana"
/mol_type="genomic DNA"
/strain="ATCC 51694"
/db_xref="taxon:803"
1..1644
/gene="groEL"
1..1644
/gene="groEL"
/notes="groEL"
/codon_start=1
/translation="MAAKEVKFGREARERLLRGVDILANAVKVLGPKGRNVVDKSF
GAPRTKDGVSVAKEIELEDFENMGAKMLREVASKTNDIAGDTTATVILGQAVOE
GVKAAAGNMPMDLKRIGDAAEVVEVGNLFPKAKIKTSAEIAQVGTISANGAAEIGK
MTADAMEKVGNEGIVTEAKTELEVEVGMQDFRGLSPVFTVNDKRWADLDDP
YILHKKLSNGLSPVLEAVVQSGKPELLIIAEDVEGEALATLVNKLGGGKIAAV
KAPGDFRRKMLIEDIALLTSQGVISDVGKLENTVLMGLAKKVNISKENTIID
GAKKAEINARVNIQVIBETSDYDREKLQERLAKLAGGAVIRVGGATEVEVKEK
KDRDDANARVAAVEGIVAGGTALLRAANALAKGSNPDQFAGINIVRRALQAPA
ROJATNAGBEAAIIVGKLENNADTFYNTATGPGDLIALGIIVDPVKVRSALQNA
SIASLITTEAWAEVPEKDTMPMPMEGGMGGMGMDP"
BASE COUNT 478 a 298 c 433 g 435 t
ORIGIN
Query Match 46.3%; Score 762.6; DB 1; Length 1644;
Best Local Similarity 67.8%; Pred. No. 7.1e-126;
Matches 1113; Conservative 0; Mismatches 519; Indels 9; Gaps 3;
Qy 1 ATGCGTCTTAAAGAAATCTCTTTTGTATGCTGTAAGCCGCTGAAAACTTTCACGAGGTGA 60
Db 1 ATGCGTCTTAAAGAAATCTCTCTGCGCGGTGAAGCGGTGAGCGTCTTTGTTGCGCGCGGT 60
Qy 61 GATAAATCTTCAATGCTGTTAAAGTAACACTTGAACCTAAAGCCGCTAATGCTGTTAT 120
Db 61 GATAATCTTCTTACGCTGTTAAAGTAACGCTCGCCCTTAAGTCGCAATGTTGTAAT 120
Qy 121 GAAAGTCTTTTGGTCTCCCGAGTTTATACAAAAGATGCTGATCTGTTGCCAAAAAGAAAT 180
Db 121 GATAAATCTTGTGCGCCTCGCATCACAAAAGATGCTGATCTGTTGCCAAAAAGAAATC 180
Qy 181 GAATCTGAAGATAGTTTGAATAATAGGCGCTCAATAGTTTAAAGAGTAGCTCCCAA 240
Db 181 GAATCTGAAGATAGTTTGAATAATAGGCGCTCAATAGTTTAAAGAGTAGCTCCCAA 240
Qy 241 ACTAGCATATTCGCTGATGGAACCTACACACACAGCTCCTTGCACAGCTATTAT 300
Db 241 ACCAATGACATTCGCTGGGATGGAACACCACTCCAATGTTTGGACAGGCTATTGTA 300
Qy 301 CGTGAAGCTGTAAACCTGTAGCAGCTGCTGTAATCCTATGGCCATTAAACGCTGCATA 360
```

301 CAAGAAGGGTGAAGCCGCTGCTGTCATGAACCCCAATGATCTCAACCGGGATT 360
361 GATAAAGCTGTGTGCTGTTACTAAAGAACTAAGCGCATTACAAAGCCTACTCGTGAC 420
361 GATGCTGCTGTGAAGAAGTGTGGTAACTCTTTTCAAAAAGCGAAAAATCCAGACT 420
421 CAAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAAATAGGTAAT 480
421 TCAGCAGAAATTCGCAAGTGGGAAACAATTCGGCCAATGGTCTGCAGAAATCGGTAA 480
481 ATCATAGCTGAAGCTGCTTAAAGTTGAAAAGAGGTGTTATCACAGTTGAGGAGCT 540
481 ATGATCGCTGATGCTATGAAAAGTGTGCAATGAAGGCTTATCACCGTGAAGAAGCG 540
541 AAAGGCTCTTGAACACTACATTAGATGTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
541 AAAAGCTGCTGAACGAAATGGAAGTCGTGGAAGGATGCAATTCGACCGTGATATCTT 600
601 TCTCATATCTTGTAACTAATCTCTGAGAAATGGTTTGTGAACCTTGATAAACCTTATATC 660
601 TCCCTCTATTTCTGTCACCAATGCTGATAAATGGTGGCTGATCTCGACGATCCTTACATC 660
661 CTTTGTATGAGAAAAGATTACTAGCATGAAGCATGCTACCAATCTTAGAACAGTT 720
661 CTCATTCAGAAAAGAAATGTCTAATTTGCAATCTTCTGCTTCCAGTGTCTGAAAGCTGT 720
721 GCTAAGTAAACCGTCCACTCTCTTATTATGCTGAAGACGTAGAAGGTGAAGCACTTGA 780
721 GTTCACTGGAACACTCTCTTATTTATCTGCTGAAGATGTGAAGGTGAGGCTTGGCA 780
781 ACATTTGATGCAATTAAGCTCGTGAGCACTTCAAGTTGTAGCCGTAAAGCTCTCTGT 840
781 AGCTCGTTTGAACAGCTGCGCGTGTGTTGAAAATGCTGCTGTGAAGCTCCAGGA 840
841 TTTGCTGAACCGCTTAAAGCTATGCTGAGATATGCTATCTTCTACTGAGAGAGCA 900
841 TTTGCTGATCGCGTGAAGCCAGTTGGAAGATATGCAATCTTGACATCGGCTCAGGT 900
901 ATATTGAAGATCGTGTGATTAAGCTTGAAGCTTGAAGCTTCTTCTTTAGGAACAGCT 960
901 ATTTCTGAAGATGTTGGCATTAATTAAGATATGTTTGTGATGCTTGTGCTGGG 960
961 AAACGTGTATTGTAAGAAATATCTATCTGTTGATGTTGCTGGAATAATCAGAA 1020
961 AAGAAGTGAATTTCTAAGAAATATACGACATTTATGATGCTGCTGGAATAATGCT 1020
1021 GATATTAAGCTCGAGTTAAACAATTCGTGACAAATTTGAGAAACAGCTCAGATTAT 1080
1021 GAAATTAATGACCGTTAAACCAATCAAGGTACAGATCGAAGAAACAACTTCTGACTAT 1080
1081 GATCGTGAACACTTCAAGAACGCTCTGCAAACTTGTGTTGGAAGTACGTTATCCAT 1140
1081 GACCGTGAACAAATTCAGAAAGACTTGTAACTCGTGGTGTGTTGCTGTTATCGT 1140
1141 GTTGAAGCTGCTACTGAACCTGAATGAAGAAAGAGATGCTGTGAGAAAGATCCTTA 1200
1141 GTTGTGTGTCGACAGAAATGTAAGTGAAGAAAGAAAGAGATGCTGTTGATGCTGCTG 1200
1201 AATCAACAGAGCTGCGGTTGAAGAGTATGTCCTGCTGCTGCTGCTGCTGCTTTC 1260
1201 AATCAACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
1261 CGCTCCATTAAGCTCTTGTATGATATTAACCTGCTGATGATGATGAACCTGCTGCTG 1320
1261 CGTGTGCAATGCTCTT--GGCAATTAAGGAAGCAATCTGATCAAGAGCTGCTATC 1317
1321 AATATCATCGCTGCTCTTGAAGAGCTTTACGTCAAAATGCTGCAAAATGCTGCTGCTAT 1380
1318 AATATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
1381 GAAGCTTCTATTGTTGTAAGAAAGCTTCTGTAACCAAA--AGATGGTTTGGATTAAT 1437
1378 GAAGCAGCAATATCTGTTGGCAAGTGTCTGAAATAACCGCAGATACATTTGGTTACAA 1437

1438 GTTGCATCAGAGAAATATGAAGACCTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCT 1497
1438 ACGCACTGCTCAATTTGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1497
1498 ACACGTATTGCAATTAACAAATCAGCATCAGTACCTTCTTCTTCTTCTTCTTCTTCTTCT 1557
1498 GTGCGTTCTGCTCTCCAGAAATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1557
1558 GCTATTGCTGAAACACCAACCACTTAAAGAAAGATATGCTCT---ATGCTGCGGTGCTGATG 1614
1558 ATGTTGCTGAAATGCTTCAAGAAAGACACTCCAATGCTCTCAATGCTGCTGCTGCTGCTGCT 1617
1615 GGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1635
1618 GGTGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1638

RESULT 11

LPNHTPB

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1100; Conservative

45.9%; Score 756; DB 1; Length 2016;

67.3%; Pred. No. 1e-124;

0; Mismatches 525; Indels 9; Gaps 2;

LPNHTPB 2016 bp DNA linear BCT 26-APR-1993
Legionella pneumophila 58-kDa common antigen (htpB) gene, complete cds.

M91673.1 M35149
M91673.1 GI:149689
58-kDa common antigen; htpB gene.
Legionella pneumophila
Legionella pneumophila

Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
Legionellaceae; Legionella.
1 (bases 1 to 2016)

Sampson, J. S., O'Connor, S. P., Holloway, B. P., Plikaytis, B. B.,
Carlone, G. M., and Mayer, L. W.,
Nucleotide sequence of htpB, the Legionella pneumophila gene
encoding the 58-kilodalton (kDa) common antigen, formerly
designated the 60-kDa common antigen

Infect. Immun. 58 (9), 3154-3157 (1990)
90354095
2117582

Original source text: Legionella pneumophila (strain Philadelphia
1, sub-species pneumophila) DNA.
Location/Qualifiers
1..2016

/organism="Legionella pneumophila"
/mol_type="genomic DNA"
/strain="Philadelphia 1"
/sub_species="pneumophila"
/db_xref="taxon:446"
211..1854
/gene="htpB"
211..1854
/gene="htpB"
/codon_start=1
/transl_table=11
/product="58-kDa common antigen"
/protein_id="AAA25298.1"
/db_xref="GI:149690"

/translation="MAKELRGDDARLQMLAGVVALADAVQVTMGPRGNVLEKSYG
APT VKGQSVAKETEFEPHFMNCAQMKVEASTDGTAGDTTATVTLARSILVEG
HVAAGMNPMDLRGDKAVLAVTKLQAMSKPKCKSKAIQVGTTSANSDAIGAI
IAEMEKVGEVITVEDGNELSVGEVQLIAVHSPYFINNQNMCELEHPI
LLVKKYSSIREMLSVLEGVAKSGRPILLIAEDVEGEALATLVNNMRGIKVCVAKA
PGFDRKMLQDAILTKGQVIFEEIGKSLLEGATLEDLSAKRIVTKENTTILDE
GKATEINARITQIEAOMEETSDYDRKLQERVAKLGGAVIKVGAATEVEMEKKA
RVEDALHTRAEVGGVAGGVALIRAKALSLKNDNDQNMGINLIRRALESFMR
QIVTNAGYASVVVNAHKDNYGNAATGEYGDWEMGLDPTKTVTRMALONAAVS
ASMLITTECMVADLPKKEBVGAGDMGGMGMGMGMGM"

591 a 323 c 508 g 594 t

BASE COUNT 591 a 323 c 508 g 594 t

ORIGIN

Query Match 45.9%; Score 756; DB 1; Length 2016;

Best Local Similarity 67.3%; Pred. No. 1e-124;

Matches 1100; Conservative 0; Mismatches 525; Indels 9; Gaps 2;

QY 8 CTAAGAAATCCTTTTGTGATGCTAAAGCCGCTGAAGAACTTTTCACGAGCTGTAGTAAC 67
DB 215 CTAAGAAATACGTTTGTGATGATGAGCTCGCTCAAAATGCTTGTGTTTAAATGCAAT 274
QY 68 TTGCAAACTCCTTTAAAGTAACACTTGGACCTAAAGCCGCTAAATGCTGTTATTTGAAAAGT 127
DB 275 TAGCAGATCGGTTCAAGTTACTATGCGTCCAGTGGTCTGTAATGTTGTTATTTGGAATAAT 334
QY 128 CTTTGGTTCCTCCAGTTATTAACAAAGATGGTGTATCTGTTCGCAAAAGAAATTTGAACCTTG 187
DB 335 CTTATGGCGCTCCTACTGCTAACTAAAGACGGTGTCTGTGTGCAAAAGAAATTTGAGTTTG 394
QY 188 AGATAAGTTTGAATATGCGGCTCAATGTTTAAAGAGTAGTCTCCCAAACTAGCG 247
DB 395 AGCATGTTTTCATGAACATGGCGCTCAAAATGTTTAAAGAGTGGTCTCTAAAACTTCG 454
QY 248 ATATTGCTGCTGATGAACCTACAAACAGCAACAGTCTCTGCAACAGCTATTTATTCGTGAAG 307
DB 455 ATACTGCTGCTGATGTACTACTACTGCAACAGTATTTGCTCGTTCTATTTCTTTGTTGAAG 514
QY 308 GTGTAAACCTTGTAGCAGTGTGCTGTAATCTATGCGCATTTAAACGTGGCATAGATAAG 367
DB 515 GTCAACAAAGCAGTTGCTGCTGTTATGAATCCAAATGGATCTCAAAACGCGTATTTGATAAG 574
QY 368 CTGTTGCTGTTTACTAAGAACTAAGCAGCAATTAACAGCCTACTCTGTCACCAAAAG 427
DB 575 CAGTATTAGCAGTTACCAAAATTAACAGCTATGCTAAGCCATGCAAAAGCAGCAAAAG 634
QY 428 AATAGCTCAAGTTGGAACCATTTCTGCAACTCTGATACAAATAGGTAAATCATAG 487
DB 635 CTAATGCTCMAAGTTGGAACCTATTTCTGCTATTTCCGATGAAGCGATTTGGTCTATCATTG 694
QY 488 CTGAAGCTATGCTAAAGTTGGAAGAGGAGTGTATTCACAGTTGAGGAAGCTTAAAGTTC 547
DB 695 CTGAAGCAATGGAAGAAAGTTGGTAAAGAGGTTGTTATACCGTTGAAGATGGTAAATGAT 754
QY 548 TTGAACACTACATTAGATGCTGTTGAAGGAATGAAGTTGACCGTGGCTACTCTCCAT 607
DB 755 TGGAAATGAGCTTCTGTTGTTGAAGGTATGCAATTTGATC---GCGGTACATCTCCAT 811
QY 608 ACTTTGTAACCTAATCTGAGAAATGTTTGTGAATTTGATTAACCTTTATCTCTTGTA 667
DB 812 ACTTTATCAACACAGCAAAACATGAGCTGTGAACTTTGAGCATCCATTCATTTTATGG 871
QY 668 ATGAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTTGCTAAAG 727
DB 872 TTGACAAAAGATTTCAGTATTTCGTGAATTTGTCGGTATTGGAAGGTGTTGCCAAAT 931
QY 728 TAAACCGTCCACTCTTATTATTGCTGAAGACGTAGAGGTGAAGCACTTGCACACTTG 787
DB 932 CTGGTCTGCTCTTATTGATCATTTGCAGAAGATGTTGAAGCGAAGCTTTAGCTACTCTG 991
QY 788 TAGTCAATAAGCTCCGTTGAGCACTCCCAAGTTGTAGCGCTAAAGCTCTCTGTTTGGTG 847
DB 992 TAGTCAACACATGCGCGGTATTGTAAGATGCTGTCTCAAGCCCTGTTTGGTG 1051
QY 848 AAGCCGCTGAAGCTATCTTTGAAGATATGCTATCTTACTGAGGAGAGCAATATTTG 907
DB 1052 ATCGCCGCAAGCGATGTCACAGACATGCTATTTTGTAGTGAAGGTCAAGTTATTTCTG 1111
QY 908 AAGATCGTGTATTAAGCTTGAATATGTAAGCTTGTCTTTTAGGAACAGCTTAAAGTG 967
DB 1112 AAGAAATGGAAGAGCTTGAAGGTGCTACTCTGGAAGATCTTTGAGTGTCTGAAGCGAA 1171
QY 968 TAGTTATTGAAGAAATTAAGTACTACTGTTGATGCTGCTGGAATCTGGAAGATATTA 1027
DB 1172 TCGTTGTTACCAAGAAACACTACTATCATTTGATGTTGAAGGAAGGCAACTGAATTA 1231
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCAAAATTTGAAGAAACAAAGCTCAGATTATGATGCTG 1087
DB 1232 ATGCTCGTATTACTCAAAATTCGTGCAAAATTTGAAGAAACCACTCTCTGATTACGATAG 1291

QY 1088 AAAAACTTCAAGAAAGCTCTTGCAAAACCTTGTGGTAGTAGCTGTATCCATGTTGGAG 1147
DB 1292 AAAAATTACAAGAGCGGTTGCTAAACTAGCTGGTGTGTGCTGTATCAAAAGTTGGG 1351
QY 1148 CTGCTACTGAACCTGAATGAAGAGAGAGATCGTGTAGAGATGCTCTTAATGCA 1207
DB 1352 CTGCTACAGAAGTTGAAATGAAGAGAGAGAGACAGCTGTGGAAGATGCTCTTCATGCTA 1411
QY 1208 CAAGAGCTCGGTTGAAGAGGTTATGCTCCCTGGTGTGCTACTGCTTTTGTCCGCTCCA 1267
DB 1412 CTCGCGCTCAGTAGAAGAGGTTATGTTGCCGCTGGTGTGCTTGAATTCGTC 1471
QY 1268 TTAAGCTCTTGAATATTAACCTGCTGATGATGAACCTTGTGGAATTAATCA 1327
DB 1472 AGAAGCTCTTGAATTAACCTGCTGATGATGAACCTTGTGGAATTAATCA 1531
QY 1328 TCGCTGCTCTCTTGAAGAGCTTTACGTCAAATTCCTGCAATTCGCTGCTGTAAGGTT 1387
DB 1532 TAGCTCGCTATTTGAATCTCAATGCTCAATTTGTTACTAACGAGGATATGAAGCTT 1591
QY 1388 CTATTGCTTGAAGAAAGTTTCGTGAACCAAAAGATGTTTGGATTTAATGCTGCATCAG 1447
DB 1592 CTGTTAGTAAACAAGGAGTGTGACACAAAGCAACTACGTTTCAACGCTGCACTG 1651
QY 1448 GAGATATGAAGACCTTTATTAAGCTGCTGTCATTGATCTTAAAGAAAGTTTACACGTTATG 1507
DB 1652 GTGAATACGGTGATATGTTGAGATGCTGTTTCTTGATCCAAAGTAAACCGTATGG 1711
QY 1508 CATTAACAAATGAGCATCAGTAGCTCTTACTTCTTACTACAGATGCTGCTATTCGCTG 1567
DB 1712 CTCTGCAAAATGAGCTTCTGTAGCCAGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTG 1771
QY 1568 AAAACCAAGCACTTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1627
DB 1772 A-----TCGCTTGAAGAAAGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1825
QY 1628 GTATGACGCTATG 1641
DB 1826 GAATGGTGGCATG 1839

RESULT 12
LOCUS AF014829
DEFINITION Bartonella henselae 60 kDa heat shock protein (groEL) gene, partial cds.
ACCESSION AF014829
VERSION AF014829.1 GI:3603158
KEYWORDS
SOURCE Bartonella henselae
ORGANISM Bartonella henselae
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bartonellaceae; Bartonella.
REFERENCE 1 (bases 1 to 1627)
AUTHORS Marston, E.L., Sumner, J.W. and Regnery, R.L.
TITLE Evaluation of intraspecies genetic variation within the 60 kDa heat-shock protein gene (groEL) of Bartonella species
JOURNAL Int. J. Syst. Bacteriol. 49 Pt 3, 1015-1023 (1999)
MEDLINE 99354490
PUBMED 10425758
REFERENCE 2 (bases 1 to 1627)
AUTHORS Marston, E.L.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1997) Viral and Rickettsial Zoonoses Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G13, Atlanta, GA 30333, USA
FEATURES
source Location/Qualifiers
1. .1627 /organism="Bartonella henselae"
/mol_type="genomic DNA"
/strain="Houston-1"
/db_xref="taxon:38323"
111. .>1627 gene

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Buchnera.

1
 Pares, M.A., Barrio, E., Sabater-Munoz, B. and Moya, A.
 The evolution of the heat-shock protein GroEL from Buchnera, the primary endosymbiont of aphids, is governed by positive selection
 Mol. Biol. Evol. (2002) In press
 2 (bases 1 to 1653)
 Sabater-Munoz, B.
 Direct Submission
 Submitted (12-MAR-2002) Sabater-Munoz B., Insto. Cavanilles Biodiv & Biol Evol., Universidad de Valencia, P.O. Box 2085, Valencia
 E-46071, SPAIN

FEATURES

source

1. .1653
 /organism="Buchnera aphidicola (Thelaxes suberi)"
 /mol_type="genomic DNA"
 /specific_host="Thelaxes suberi aphid"
 /db_xref="taxon:98797"
 /country="Spain:Valencia"
 1. .1653
 /gene="groEL"
 /function="molecular chaperonin"
 /codon_start=1
 /transl_table=11
 /product="GroEL protein"
 /protein_id="CAD27797.1"
 /db_xref="GI:21425627"

gene

CDS

/translation="MAAKVQFCNDARVQKMLRGVNLADAVKVLGPKGRNVLDKSF
 GAPAIYKDVTVAREIELEDKPFNMGQWKEVASKANDAGDTTATVLAQSIYVE
 GLKAAAGNMPDLKQVDKAVIAAEEELKCLSPVDFKAIQVQVGIANSDETQVK
 LIAEMKVKGEVITVEEGTQDELVDVVEGDFQXLSPIYINKPDTGSIELDNP
 FILLADKTSNNREMLPPLSVAKAGNLLIAEDVEGEALATLVNMTGIVKVAVK
 APGFGRKAMLOIATLTGSTVISEIGLEIKETITIEDIGHAKRVVINKDITTIIDG
 IGNESAIKGRVVOIROIBETSDYDREKLOERVAKLAGVAVIKVGAATEVEMKEKK
 ARVEDALHATRAVEGVAGGVGVALIRVAGKLSLITGENEDQNGVIGKVAIRAMESPL
 RQIVINAGEASAIANNVAKNGNYGNAYSEIQGDMIKMGILDPIKVRSLAQYAA
 VAGIMITECMITLPLKSSDLSNAGNGMGGMGMGMGM"

BASE COUNT 575 a 205 c 382 g 491 t

ORIGIN

Query Match

Best Local Similarity 45.7%; Score 753; DB 1; Length 1653;

Matches 1093; Conservative 66.6%; Pred. No. 3.7e-124;

1; Gaps 1; Indels 545; Mismatches 0;

QY 1 ATGCTCTTAAAGAAATCCCTTTTGTGCTAAAGCCCGTGAAAAAATTTTCACGAGGTGTA 60
 |||||
 DB 1 ATGGCAGCTAAAGACGTAAATTTGGAAATGATGCTCGCGTAAATATGCTTCGTGGAGTA 60
 |||||

QY 61 GATAAACTGCAATGCTGTTAAAGTAAACACTTGGACCTAAAGCCCGTAATGCTGTTATT 120
 |||||

DB 61 AATGCTACTGATGATGAGTAAAGTTTACTTTAGGACCAAAAGGTCGTAATGATGACTT 120
 |||||

QY 121 GAAAGTCTTTTGTGTCGCCAGTTATTACAAAGATGCTGATCTGTTGCAAGAATAATT 180
 |||||

DB 121 GATAAATCTTTTGTGTCCTGCTATTACTTAAAGATGGTGTAACTGAGCTCGTGAATC 180
 |||||

QY 181 GAACCTCAAGATAAGTTTGAATAATGCGCGCTCAATGTTTAAAGAGTAGCTCCCAAA 240
 |||||

DB 181 GAATTAGAAGATAAATTTGAACAATGGAGCTCAGATGTTAAAGAGTAGCATCTAAA 240
 |||||

QY 241 ACTAGCATATTGCTGGTGTAGGAACCTACACAGCAACAGTCCTTGCACAGCTATTATT 300
 |||||

DB 241 GCGAACGATGACGCTGGTGTAGGAACCTACACGCTACTGTTATTAGCACAATCTATTGTA 300
 |||||

QY 301 CGTGAAGGTGTAACCTTGTAGCAGCTGGTGGTAATCTCTTATGCGCATTAACAGTGGCATA 360
 |||||

DB 301 ACAGAGGGAATTAAGGCTGTGTCAGCTGGAATGAATCCAAATGGATTTAAAGAGGGGTA 360
 |||||

QY 361 GATAAAGCTGTGTTGCTGTTACTTAAGAAGCTAAGCGACATTACAAGCCCTACTCGTGAC 420
 |||||

DB 361 GATAAAGCTGTGTTATGTCAGCAGTAGAAGAACTTAAATGCTCTTCTGCTCTGTTCTGAT 420
 QY 421 CAAAGAANAATAGCTCAAGTTGCAACCATTTCTGCAAACTCTGATACAACAATAGTAAT 480
 |||||
 DB 421 TTAAAGACTATTGCTCAAGTTGCTACTATATCTGCAAAATCTGATGAAACTGTAGAAA 480
 |||||
 QY 481 ATCATAGCTGAAGCTATGCTAAAGTTGGAAGAGGAGGTGTTATCACAGTTGAGGAAGCT 540
 |||||
 DB 481 TTAATTCAGAGCANTGGAAGAGGTGGAAGAAGAGGTGTAATCACTGTTGAAGAAGGA 540
 |||||
 QY 541 AAAGGCTTTGAAACTACATTAGATGTTGGAAGGAATGAAGTTGACCGTGCTACCTC 600
 |||||
 DB 541 ACTGGAATACAGGATCAATTAGATGTTAGTAGAAGGAATGCAATTTGATCGTGTGTTACTTA 600
 |||||
 QY 601 TCTCCATACCTTTGTAACCTCTGAGAAAATGTTTGTGAACCTGATACACCTTATATC 660
 |||||
 DB 601 TCTCCTTATTTTCAATTAATAACAGATACCTGGATCTATAGAAATAGATAATCCATTTATT 660
 |||||
 QY 661 CTTTGTAAATGAGAAAAGATTTACTAGCATGAAAGACATGCTACCAATCTTTAGAACAGTT 720
 |||||
 DB 661 CTTTGTAGCAGATAAAGAGATTTCTAATAATCGAGAATGCTTCTCTCTTTAGAACTCTGTT 720
 |||||
 QY 721 GCTAAAGTAAACCGTCCACTCTCTTATTTATGCTGAGAACGCTAGAGGTGAAGCATTGCA 780
 |||||
 DB 721 GCTAAAGCA---GGAACCTTATTAATTTATGCTGAAGATGTTGAAGGTGAAGCATTTAGCA 777
 |||||
 QY 781 ACATTTGTAGTCAATAAGCTCCGTTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCCTGGT 840
 |||||
 DB 778 ACATTTAGTAGTAAATACTATGCGTGGTATAGTAAAGTAGCTCGGTAAAGCACCCTGGT 837
 |||||
 QY 841 TTTGGTGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 |||||
 DB 838 TTTGGTGAAGAGTAAAGCTATGCTACAGATATTTGCAACTTTAACAGGTAGCACAGTA 897
 |||||
 QY 901 ATATTTGAAGATCGTGGTATAAAGCTGATAAAGCTGTAAGCTTGTCTTCTTTAGGAACAGCT 960
 |||||
 DB 898 ATTTCTGAAGAAATTTGGAATTTAGAAATTTAGAAAAAATACAAATAGAAATTTAGGTACGCT 957
 |||||
 QY 961 AAAGCTGTAGTTATTGACAAAGAAATATCTACTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 |||||
 DB 958 AAAGAGTAGTGTAATTAATAAGATCTACTACTATTATTGATGCTGCTGCTGCTGCTGCTGCTGCT 1017
 |||||
 QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAGAAACAAAGCTCAGATTAT 1080
 |||||
 DB 1018 GCTATTAAGCTCGTGTGTTGCAAAATCTGCAAAATAGAGAGCAACATCTGATTAT 1077
 |||||
 QY 1081 GATCGTGAAGAACTTCAAGAACTGTTGCAAACTGTTGGTGGAGTAGCTGTTATCCAT 1140
 |||||
 DB 1078 GATCGTGAAGAAATTTACAAGAACGCTAGCTAAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
 |||||
 QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGAGGATCGGTAGAGATGCTCTA 1200
 |||||
 DB 1138 GTTGGTCTGCACTGAAATGAAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
 |||||
 QY 1201 AATGCAACAGAGCTGCGGTTGAAAGAGGATTTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 |||||
 DB 1198 CATGCAACTCGTGACGCTGTAGAGAGAGGTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 |||||
 QY 1261 CGCTCCATTAAGTCTTGTATGATTTAAACCTGCTGATGATGATGAACTTGTGCTGCTGCTGCTGCT 1320
 |||||
 DB 1258 CGTGTCTGCT 1317
 |||||
 QY 1321 AATATCATCGTCTGCTCTTGAAGACCTTTAGTCAAAATGCTGCAAAATGCTGCTGCTGCTGCTGCT 1380
 |||||
 DB 1318 AAAGTAGCTATACGTGCAATGGAATCTCCACTTCGTCAAAATAGTAAATGCAAGGTGCA 1377
 |||||
 QY 1381 GAAGGTTCTATTGTTGTAAGAAAAGTTTCGTGAACCAAAAGATGCTTGTGATTTTAAATGCT 1440
 |||||
 DB 1378 GAAGCATCTGTTATAGCAATTAATGTAAGAGCTGGAAGAGAAATACGGTTATTAATGCA 1437
 |||||
 QY 1441 GCATCAGGAGATATGAAGACCTTTATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 |||||
 DB 1438 TATTCTGAACAAATATGCTGATATGATTAAATGCTGTTATTTAGACCCCACTTAAAGTTACT 1497
 |||||

```
QY 1501 CGTATTGCAATTAACAAATCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAAATCGGCT 1560
Db 1498 AGATCTGCTTACAAATATGCTGCACTGAGTGGTTTAAATGATTAATGATTAATGATTAATG 1557
QY 1561 ATTGCTGAAGAAACAGCACTTAAGAAATATGCTATGCTGGCGGTGGTATGAGTGGT 1620
Db 1558 ATTACTGATCTACCTAAAGAGCAATTCATCAGATTTAAGCAATGCTGGAATGCTGGTATG 1617
QY 1621 ATGGGTGGTATGACGGTATG 1641
Db 1618 GGTGGTGGTATGCGGTGGCATG 1638

RESULT 14
LMGROELS
LOCUS LMGROELS 2770 bp DNA linear BCT 18-FEB-1992
DEFINITION L.micdadei DNA for gro ELS operon.
ACCESSION X57520
VERSION X57520.1 GI:44099
KEYWORDS groELS operon.
SOURCE Tatlockia micdadei
ORGANISM Tatlockia micdadei
Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
Legionellaceae; Tatlockia.
REFERENCE 1 (bases 1 to 2770)
AUTHORS Hinderson, P., Hoiby, N. and Bangsberg, J.
TITLE Sequence analysis of the Legionella micdadei groELS operon
JOURNAL FEMS Microbiol. Lett. 77, 31-38 (1991)
FEATURES
    location/Qualifiers
        1..2770
            /organism="Tatlockia micdadei"
            /mol_type="genomic DNA"
            /db_xref="taxon:451"
BASE COUNT 846 a 520 c 675 g 729 t
ORIGIN
```

Query Match 45.6%; Score 751.2; DB 1; Length 2770;
Best Local Similarity 67.6%; Pred. No. 7e-124;
Matches 1104; Conservative 0; Mismatches 518; Indels 12; Gaps 3;

```
QY 8 CTAAGAATCGTTTTCATGCTAAAGCCCGTGAAGAACTTTCACGAGTGATGAATAC 67
Db 472 CTAAGAATTAAGTTTTCGCGACGATGCTCGTCAACAAATGCTGCTGTTAAACGAT 531
QY 68 TTGCAATGCTGTTAAAGTAAACACTTGGACCTAAAGCGCGTAATCTCGTTATTGAAAGT 127
Db 532 TGGCTGATGAGTTAAAGCAATGCGGTCCAGCGGACGTAACTGTTGTTTGAAGAAAT 591
QY 128 CTTTGGTCCCGAGTTATTAAAGATGTTGATCTCTGTTGCAAGAAATTTGAACCTTG 187
Db 592 CATTTGGTGCACCTACTGTAACAAAGACGGTGTCTGTTGCAAGAAATTTGAATTCG 651
QY 188 AAGATAAGTTTGAATATGCGCTCAATGTTAAAGATGACTGCCAAATACG 247
Db 652 AAAACCGTTTCAAAAATATGCGGCACAAATGTTTAAAGATTTGCTGCTAAATCTG 711
QY 248 ATATTGCTGGTATGGAAGTAAACAGCAACAGTCTTTCGCAAGCAATTTATCGTGAAG 307
Db 712 ATACCGAGGTGATGTTACTACTGCTACGTATGCGATGCGGCTGCTGCTGTTGAAG 771
QY 308 GTGTAAGACTTGTAGCAGCTGGTGTAAATCTTATGCGCAATTTAAAGCGTCATAGTAAAG 367
Db 772 GCCCAAGACAGTTGACGAGCATGAACCAATGATTTGAAAGAGGATTTGATAAG 831
QY 368 CTGTTGTTGCTTACTTAAGAACTAAGCAGATTAAGAGCTACTCGTGACCAAAAG 427
Db 832 CAGTTGCGCGAGTTACCAAAAATTTACAAAGAAATGCTTAAGCCTTGAAGACCGTAAAG 891
QY 428 AAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGTATACAAATAGTAAATCATAG 487
Db 892 CAATTGCAACAGTAGTACTATTTCTGGAACTCTGATCAAGCAATCGTTCTATCATG 951
```

```
QY 488 CTGAAGCTATGCTAAAGTTGAAAGAGGAGTGTATACAGTTGAGGAGCTAAAGGTC 547
Db 952 CTGAGCAATGAAAGAGTAGTTAAAGAGGTTTATTAATCTGTAAGATGTTAAAGCC 1011
QY 548 TTGAACACTATAGATGTTGTAAGGAATGAAGTTTCAAGTGGCTACTCTCTCCAT 607
Db 1012 TTGAACCAATATAGCTGTTGTTGAAGGTATGAGTTCATGCTGTTTCAATTTCTCCAT 1071
QY 608 ACTTTGTAATAATCCTGAGAAATGTTTGTGAAGTTCATTAACCTTATATCTTTGTA 667
Db 1072 ACTTCATCAACATCAGCAAAATCATGAGTGTGAGCTTCAGCATCTCTTCCTCTGG 1131
QY 668 ATGAGAAAGATTTACTAGCATGAAGACATGCTACCAATCTTAGAAACAGTTGCTAAAG 727
Db 1132 TTGACAGAAATTTCACAATTCGTGACATGTTGCTGTATTGAGAGCAGTTGCCAAT 1191
QY 728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGTGAAGCACTTGAACACTTG 787
Db 1192 CAGGACCGTCTTTTATTGATCATTTGCTGAAGATGTTGAAGGTGAAGCATTAGCAACTCTG 1251
QY 788 TAGTCAATAAGCTCCGTGAGCACTCCCAAGTTTGTAGCCGTAAAGCTCCTGTTTGGTG 847
Db 1252 TTGTTAAACAATGCGCGGTATTGTTAAAGTGTGCGCGTTAAAGCACTGTTTCCGTTG 1311
QY 848 AACGCGCTAAAGCTATGCTTGAAGATATTGCTATCTTACTGAGGAGAGCAATATTG 907
Db 1312 ATCGTGTGAAGCCATGTTACAGACATTTGCTATCTTACTGCGGCGAGTCATTTTCAG 1371
QY 908 AAGATCGGTGATATAAGCTTGAAGTTTGAAGTCTGCTTCTTTTAGGAACAGCTAAACGTG 967
Db 1372 AAGAAGTGGTACTAGTTTAGAAAGTGTACTTTAGACAGCTGAGCACTGCAAAACGG 1431
QY 968 TAGTTATTACAAAGAAATATCTACTATCTGTTGATGGTGTCTGGAATCAGAAATATTA 1027
Db 1432 TTGCTGTACCAAGAAACACCAATCATTTGACGGTGAAGGTAAAGCGGTGACATCA 1491
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTCAGAAACAAAGCTCAGATTTATGTCGTG 1087
Db 1492 ATGCGGTATTACTCAAAATTCGTGCTCAATGGAAGAAACCACTTCTGATTAAGTCCGG 1551
QY 1088 AAAAATTTCAAGAACTTTGCAAACTTTGTTGGTGGAGTACTGTTATCCATGTTGGAG 1147
Db 1552 AAAAATTTCAAGAGCGCGTAGCTAAATTAGCTGGCGGTGTTGCCGTCTAT-TAAGTTGGTG 1610
QY 1148 CTGCTACTGAACTCAAAATGAAGAGAGAGAGGATCTGTAGAAAGTCTCTTAATGCA 1207
Db 1611 CTGCTACCGAATC-----GAATGAAGAGAAAGCTCGTGTAGAGACGCTCTTCATGCAA 1665
QY 1208 CAAGAGCTGCGGTTGAAGAAAGTATTGTCCTGTTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db 1666 CTCGTGCTGAGTAGAAGAGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1725
QY 1268 TTAAGTCTTGTATGATTAATTAACCTGCTGATGATGATGAATCTTGTGCTGACTTAATCA 1327
Db 1726 AAAAAGCATTAGATGGCTTGAAGGCGAAATGCTGACCAAGACATGGGTATTAAATCC 1785
QY 1328 TCCGTGCTTCTTGAAGAGCTTTTACCTCAATTTGCTGCAAACTGCTGCTATGAGGTT 1387
Db 1786 TACGTGCGCGCATCGAATCTCTCTCCCGCAAACTGATAGCAATGATGAGGCTTATGAGCCTT 1845
QY 1388 CTTATGTTGTAAGAAAGTTTCGTGAACCAAGAGTGGTTTGGATTTTAAATGCTGCTCAG 1447
Db 1846 CTGTTATCGTTAACAAAGTAGCTGAAGCAAGATAACTTTGGTTTCAACGACGACTCTG 1905
QY 1448 GAGAAATGAAGACCTTATTAAAGCTGGTGTCAATTGATCTTAAAGGTTTACAGTATG 1507
Db 1906 GCGAATACGCTGACATGGTTTGAAGATGGGTATTCTTGTATCCAACTAAAGTACTCTACTG 1965
QY 1508 CATTACAAATGACAGCATCAGTACCTCTTACTTCTAATCTACAGAACTGCTATGCTG 1567
Db 1966 CATTACAAAGCGAGCTTCTGTTGCAAGCTTAATGCTGACCAAGTGGTGGTGGTGGTGG 2025
QY 1568 AAAAACCAAGACCTAAAGAAAGATATGCTATGCTGCGCGGTGGTATGCGGTGGTATGGGTG 1627
```

Db 2026 A-----TCTACTAAGAAGAGAACCTATGGTCCCGGCAATGGCGCATGGGTG 2079
QY 1628 GTATGACGGTATG 1641
Db 2080 GAATGGCGGCATG 2093

RESULT 15
AX607355
LOCUS 1623 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 5284 from Patent WO20292818.
ACCESSION AX607355
VERSION AX607355.1 GI:28402886
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Glaser, P., Rusniok, C., Chevalier, F., Frangeul, L., Lalioui, L.,
Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Trieu-Cuot, P. and
Kunst, F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 5284 21-NOV-2002;
INSTITUT PASTEUR (IPR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (IPR)
FEATURES
source
1. .1623
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/db_xref="taxon:1311"
BASE COUNT 499 a 271 c 391 g 462 t
ORIGIN
Query Match 45.3%; Score 746.6; DB 6; Length 1623;
Best Local Similarity 68.0%; Pred. No. 5e-123;
Matches 1072; Conservative 0; Mismatches 499; Indels 6; Gaps 2;

QY 8 CTAAAGAAATCCCTTTTGTATGCTTAAGCCCGTGAAACCTTTTACAGAGGTGTAGATAAAC 67
Db 5 CAAGAATATTAATTTTACAGAGATGCCCGCTCAGCAATGGTGGGTGTGATATTT 64
QY 68 TTGCAATGCTGTTAAAGTAACCTTGGACCTAAAGCCCGTAATGCTGTTATTTGAAAGT 127
Db 65 TAGCTGATACAGTCAAAAGTAACATTAAGTCTTAAAGCCCGTAATGTTGTTCTTTGAAAG 124
QY 128 CTTTTGTGTTCCCGAGTTATTACAAAGATGCTGTATCTGTTGCAAAAGAAATGAACCTG 187
Db 125 CATTTGGTTCTCTTTAAATTAACAATGATGGTGTGACAATGCTTAAGAAATTTGAGCTAG 184
QY 188 AAGATAAGTTTGAATPATGGGCGCTCAAAATGGTTAAAGAGTAGTCTCCCAAACTAGCG 247
Db 185 AAGATCACTTTGAAATATGGAGCTAAACTTGTGTCAGAAAGTGGCTTCAAAACTAATG 244
QY 248 ATATTGCTGTGATGAACTACAAACAGAACAGTCCTTGGCAACAGCTATTATTGCTGAAG 307
Db 245 ATATTGAGGGGATGGCACTACAACTGCTACTGTTTGAACCCAGCTATTGTACGGGAAG 304
QY 308 GTGTAAACCTTTGAGAGCTGGTGTGTAATCCTATGGCCATTAAACGGGATAGATAAAG 367
Db 305 GTCTTAAATATGTAACCTGAGGGGCAATCCGATGGCAATTCGTGGTATTTGAAACAG 364
QY 368 CTGTTGTTGCTGTTTACTAAAGAACTAAGCGCACTTACAAAGCCCTACTCGTGACCAAAAG 427
Db 365 CTGTTTACAGCAGCTGTTGAAGAGCTAAGAGAGATTGCAACACCACTTTCAGGCAAGAG 424
QY 428 AATAGCTCAAGTTGGAAACCAATTTCTGAACTCTGATACAACTAGTAATATCATAG 487
Db 425 CTATTGCTCAAGTTGCGGCTGTGCTTTCACGTTCTGA---AAAAGTTGGGGAATACATTT 481
QY 488 CTGAGCTATGGCTAAAGTTGGAAGAGGAGGTGTTATACAGTTGAGGAAGCTAAAGTC 547

Db 482 CTGAGCTATGGGCGCGTGGGTAAATGATGGTGTATCACTATTGAAGAATCGGAGTA 541
QY 548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGCTACCTCTCTCCAT 607
Db 542 TGAAGAACAGAGCTTGAAGTTGTGAAGGAATGAGTTTGACCGTGGGTACTTGTACAGT 601
QY 608 ACTTTGTAATACTCTGAGAAATGGTTTGTGAACCTTGATACACCTTATATCTTTTGT 667
Db 602 ATATGTAATACTGATAACGAGAAATGGTCTCTGAACCTTGAGATCGTATATCTTATTA 661
QY 668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTTAGAACAAAGTTGCTAAAG 727
Db 662 CAGATAAGAAAAATTTCAAATATCCAAGAAATTTTACCAATTTATTAGAAGAGTTCTTAAAA 721
QY 728 TAAACGCTCACCTCTTATTTATTTGCTGAAGACGTAGAAGGTGAAGCACITGGAACACTTG 787
Db 722 CAAATCGCTCGTTGCTTAATCATCGCTGATGATGTTGATGAGAAAGCTCTCCCAACGCTG 781
QY 788 TAGTCAATAAGCTCCGCTGGAGCACTCCAAGTTTGTAGCGTAAAGCTCTCTGGTTTGGTG 847
Db 782 TTCTTAAACAAATTCGTGGAACCTTCAATGTCGACTGTAAAGCGCTGATTTGGTG 841
QY 848 AACCGCTTAAAGCTATGCTTGAAGATATTCTTATCTTACTTGAGGAGAAAGCAATATTG 907
Db 842 ATCGTCGTAAGCCATGCTGGAAGATATTGCTATCTTAACAGGAGAACTGTCTGTACTG 901
QY 908 AGATCGTGTATAAAGCTTGAAGATTAAGCTTGTCTTCTTTAGGAACAGCTTAAACGCTG 967
Db 902 AAGACCTTGGTTTAGACTTTAAAGATGCTACTATGCAAGTTTGTAGGACAGCTCTGCTAAAG 961
QY 968 TAGTTATTGACAAAGAAATTAATCTATCTGCTGCTGCAAAATCAGAAAGATTA 1027
Db 962 TAAAGTAGTAAGAATTCTACTGTTATTTGTCGAGGTGCGGTGACTCATCAGCAATTG 1021
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCAAAATTTGAAGAAACAAAGCTCAGATTATGATCGTG 1087
Db 1022 CTAATCGGTAGCTATCATTAAGTCACAGATGAGGCTACAACCTTCTGATTTTGTATCGTG 1081
QY 1088 AAAAATTCAAGAACGCTTTGCAAACTTGTGCTGAGTAGCTGTTATCCATGTTGGAG 1147
Db 1082 AAAAATTCAAGAACGCTTTGCTAAGTTAGCCGCTGCTGAGCAATTAATTAAGTTGGTG 1141
QY 1148 CTGCTACTGAACTGAAATGAAAGAGAGAGATCGTGTAGAGATGCTCTAAATGCA 1207
Db 1142 CAGGACTGAAACAGAAATTAAGAGAGATGAACCTTGCATCGAGATGCGTTAAATGCA 1201
QY 1208 CAAGAGCTCGGTTGAAGAGAGTATTGTCCTCGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db 1202 CGGTCGCTCAGTTGAAGAGAGTATTGTTTCAGGTGGAGTACGGCTCTTGTGAACGTTA 1261
QY 1268 TTAAGTCTTGTATGATATTAAACCTGCTGATGATGATGAACCTTGTGGACTTAATATCA 1327
Db 1262 TTGAAAAGTGAAGGAGCTGAAACCTTAAATGCTGATGAGGA---GACTGGACGTAATATG 1318
QY 1328 TCCTGCTCTCTCTTGAAGAGCCCTTACGTCAAATTCGTCAAATTCGTGCTGCTTATGAAGTT 1387
Db 1319 TTTCTGCTGCTCTGGAAGAGCCCTTGTGCTCAAATTCGTCAAATTCGTGCTGATGAGAGTT 1378
QY 1388 CTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGTTTGTGATTTTAAATGCTCATCAG 1447
Db 1379 CAGTTATTATTGAACGTTTAAACAGCTCTGAAATTTGTTACAGGATTTAATGCGGCCAATG 1438
QY 1448 GAGATATGAAGACCTTATTAAAGCTGGTGTCTATGATCTCTAAAGAAAGTTTACAGTTATG 1507
Db 1439 GAGAAATGGGTAGATATGTTTACCACAGGTATCATTTGACCCCTGTCAAAGTAAACAGTTCTG 1498
QY 1508 CATTTACAAATGAGCAATCAGTACGCTCTCTTACTTCTAACTACAGAAATGCGCTATTCGTG 1567
Db 1499 CACTTCAAAATGGGCACTCTGTAGCAAGCTTTATCTTCACTACAGAAAGCTAGTAGCA 1558
QY 1568 AAAAACCAAGAACCTTAA 1584

Db 1559 ATAAACCTGAACGAGAA 1575

Search completed: January 29, 2004, 14:44:06
Job time : 4112 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 09:58:37 ; Search time 352 Seconds

(without alignments)
12630.609 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 1647

Sequence: 1 agggcttcaagaatactct.....gtatggacggtatgtag 1647

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 19Jun03:*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	1647	18 AAT69201	Lawsonia intracell
2	1647	100.0	4972	18 AAT69203	Lawsonia intracell
3	746.6	45.3	1620	24 ABN69159	Streptococcus poly
4	746.6	45.3	2155561	24 ABN71527	Streptococcus poly
5	741.8	45.0	96109	22 AAF28548	Genomic fragment #
6	732.2	44.5	1901	24 ABQ70705	Listeria monocytog
7	729	44.3	213251	24 ABQ67193	Listeria innocua c
8	729	44.3	3011208	24 ABQ69245	Listeria innocua D

c	9	722.6	43.9	2944528	24	ABA03041	Listeria monocytoc
	10	721	43.8	1830121	17	AAT42063	Haemophilus influe
	11	713.6	43.3	2465	12	AAQ13136	Hyp operon. Chlam
	12	710.4	43.1	2223	12	AAQ13137	Hyp operon. Chlam
	13	710	43.1	640681	24	ABA92787	Buchnera sp. genom
	14	708.4	43.0	2365589	24	ABA90521	Genomic sequence o
	15	704.2	42.8	1661	20	AAH86155	DNA encoding a Str
	16	701	42.6	1635	24	ABN69160	Streptococcus poly
	17	698.4	42.4	3625	20	AAH12979	Enterococcus faeca
	18	698.4	42.4	3625	24	ABS98774	Enterococcus faeca
c	19	690.2	41.9	1230025	20	AAH91990	Nucleotide sequenc
	20	689.8	41.9	1654	20	AAH86153	DNA encoding a Str
	21	689.8	41.9	1926	22	AAF25036	Nucleotide sequenc
	22	689.6	41.9	1635	24	ABL92607	Chlamydia pneumoni
	23	689.2	41.8	2107	22	AAH56860	Antibiotic resista
	24	687.6	41.7	1620	25	ABX07694	S. pneumoniae type
	25	687.6	41.7	1647	22	AAH90906	2CFE 104 coding se
	26	687.6	41.7	5365	19	AAV52210	Streptococcus pneu
c	27	687.6	41.7	2162598	25	ABS56454	CFE 104 coding seq
	28	686	41.7	1623	22	AAH90800	Complete genome se
	29	686	41.7	1038602	20	AAZ01425	Escherichia coli G
	30	684.4	41.6	1647	22	AAH90805	E. coli groESL gen
	31	684.4	41.6	2155	24	AAH33423	E. coli groEL codi
	32	682.8	41.5	1647	21	AAA48498	Nucleotide sequenc
	33	682.8	41.5	1647	24	ABL96324	E. coli K12 strain
	34	682.8	41.5	1741	22	AAH56858	Multidrug resistan
	35	682.8	41.5	2032	22	AAH56856	Escherichia coli g
	36	682.8	41.5	4524	21	AAA48500	S. pneumoniae groE
	37	679.6	41.3	2401	22	AAH56859	Listeria monocytog
	38	675.2	41.0	1570	24	ABQ69121	Detrimental organi
	39	673.2	40.9	1647	24	ABQ81662	Detrimental organi
	40	671.6	40.8	1647	24	ABQ81665	Detrimental organi
	41	670	40.7	1647	24	ABQ81654	Detrimental organi
	42	670	40.7	1647	24	ABQ81663	Detrimental organi
	43	670	40.7	1647	24	ABQ81664	Detrimental organi
	44	667	40.5	2006	22	AAH56857	Uropathogen E. coli
	45	656	39.8	1638	24	ABQ81658	Detrimental organi

ALIGNMENTS

RESULT 1
AAT69201
ID AAT69201 standard; DNA; 1647 BP.
XX
AC AAT69201;
XX
DT 20-AUG-1997 (first entry)
XX
DE Lawsonia intracellularis GroEL DNA.
XX
KW Intestinal disease; porcine proliferative enteropathy; vaccine;
KW GroEL; heat shock protein; ss.
XX
OS Lawsonia intracellularis.
XX
EN WO9720050-A1.
XX
PD 05-JUN-1997.
XX
PF 29-NOV-1996; 96WO-AU00767.
XX
PR 30-NOV-1995; 95AU-0006911.
PR 30-NOV-1995; 95AU-0006910.
XX
(DARA-) DARATECH PTY LTD.
XX (PIGR-) PIG RES & DEV CORP.
XX
PI Hasse D, Panaccio M;
XX
DR WPI; 1997-310605/28.
DR P-PSDB; AAW16678.

XX Vaccine for treating or preventing Lawsonia intracellularis
PT infection - especially in pigs, containing non-pathogenic form of
PT bacterium or its components
XX
PS Claim 12; Page 38-42; 94pp; English.
XX
CC A DNA molecule (AAT69201) codes for the GroEL heat shock protein
CC (AAW16678) of Lawsonia intracellularis, the causative agent of porcine
CC proliferative enteropathy (PPE). A genomic library was prep'd. from
CC L. intracellularis PPE lesion isolates and screened with rabbit
CC anti-L. intracellularis antiserum. Phagemid DNA was isolated from
CC individual clones and sequenced. GroEL and GroES (see also AAT69202)
CC sequences were identified and another 13 clones were subsequently
CC (AAT69203-15) isolated. These nucleic acids can be used to produce
CC recombinant polypeptides useful in vaccines against intestinal
CC diseases such as PPE, including recombinant vaccines utilizing
CC bacterial, fungal or viral vectors. They can also be used as
CC genetic vaccines and in diagnostic assays.
XX
SQ Sequence 1647 BP; 536 A; 277 C; 368 G; 466 T; 0 other;
Query Match 100.0%; Score 1647; DB 18; Length 1647;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTCTAAAGAAATCCTTTTGTATGCTAAAGCCGTAAGAACTTTCAGAGGTGA 60
DB 1 ATGGCTTCTAAAGAAATCCTTTTGTATGCTAAAGCCGTAAGAACTTTCAGAGGTGA 60
QY 61 GATAAATCTGCAATGCTTTTAAAGTAACACTTGGACCTAAAGCCGTAATGCTGTTAT 120
DB 61 GATAAATCTGCAATGCTTTTAAAGTAACACTTGGACCTAAAGCCGTAATGCTGTTAT 120
QY 121 GAAAGTCTTTTGGTCTCCAGTATTACAAAGATGGTGTATCTGTTCAAAAGAAAT 180
DB 121 GAAAGTCTTTTGGTCTCCAGTATTACAAAGATGGTGTATCTGTTCAAAAGAAAT 180
QY 181 GAACTTGAAGATAAGTTTCAAAATATGGCGCTCAAAATGTTAAAGAGTAGTCCCAA 240
DB 181 GAACTTGAAGATAAGTTTCAAAATATGGCGCTCAAAATGTTAAAGAGTAGTCCCAA 240
QY 241 ACTAGCGATATTGCTGGTATGGAATCAACAGCAACAGTCTCTTGCAACAGTATTTAT 300
DB 241 ACTAGCGATATTGCTGGTATGGAATCAACAGCAACAGTCTCTTGCAACAGTATTTAT 300
QY 301 CGTGAAGTGTAAACTGTACAGTGTCTGTATCTTATGCGCATTTAAACGTGSCATA 360
DB 301 CGTGAAGTGTAAACTGTACAGTGTCTGTATCTTATGCGCATTTAAACGTGSCATA 360
QY 361 GATAAAGCTGTGTTGCTTACTTAAAGAACTAAGCGCATTTACAAAGCTACTCGTGAC 420
DB 361 GATAAAGCTGTGTTGCTTACTTAAAGAACTAAGCGCATTTACAAAGCTACTCGTGAC 420
QY 421 CAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGTATACAAATAGGTAAT 480
DB 421 CAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGTATACAAATAGGTAAT 480
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTGTGAAAGAGGTGTTATCACAGTTGAGGAAGCT 540
DB 481 ATCATAGCTGAAGCTATGCTTAAAGTGTGAAAGAGGTGTTATCACAGTTGAGGAAGCT 540
QY 541 AAAGTCTTTGAAACTACATTAGATGTTGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
DB 541 AAAGTCTTTGAAACTACATTAGATGTTGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
QY 601 TCTCCATATCTTGTAACTAATCTGAGAAATGGTTTGTGAACCTTGTAAACCTTTATATC 660
DB 601 TCTCCATATCTTGTAACTAATCTGAGAAATGGTTTGTGAACCTTGTAAACCTTTATATC 660
QY 661 CTTTGTATGAGAAAGAAATTAAGTATGATGAAGCATGCTACCAATCTTAGAACAGTT 720
DB 661 CTTTGTATGAGAAAGAAATTAAGTATGATGAAGCATGCTACCAATCTTAGAACAGTT 720

RESULT 2

AAT69203

ID AAT69203 standard; DNA; 4972 BP.

XX AAT69203;

XX AC AAT69203;

DT 20-AUG-1997 (first entry)

QY 721 GCTTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAAGGTGAAGCACTTGCA 780
DB 721 GCTTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAAGGTGAAGCACTTGCA 780
QY 781 ACACCTTGTAGTCAATTAAGCTCCGTGGAGCACTCAAGTTGTAGCCGTAAAGCTCCTGGT 840
DB 781 ACACCTTGTAGTCAATTAAGCTCCGTGGAGCACTCAAGTTGTAGCCGTAAAGCTCCTGGT 840
QY 841 TTTGGTGAACGGCGTAAAGCTATGCTTGAAGATATTGCTTATCTTACTGGAGGAGAAGCA 900
DB 841 TTTGGTGAACGGCGTAAAGCTATGCTTGAAGATATTGCTTATCTTACTGGAGGAGAAGCA 900
QY 901 ATATTTGAAGATCGTGTATAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 960
DB 901 ATATTTGAAGATCGTGTATAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 960
QY 961 AAACGTGTAGTTATTGACAAAGAAATACTTACTATCGTTGATGGTCTGAAATCAGAA 1020
DB 961 AAACGTGTAGTTATTGACAAAGAAATACTTACTATCGTTGATGGTCTGAAATCAGAA 1020
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTTGAAGAAACAAGCTCAGATTAT 1080
DB 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTTGAAGAAACAAGCTCAGATTAT 1080
QY 1081 GATCGTGAAGAACTTCAAGAACTCTTTCGCAAACTTGTGCGAGTAGCTGTTATCCAT 1140
DB 1081 GATCGTGAAGAACTTCAAGAACTCTTTCGCAAACTTGTGCGAGTAGCTGTTATCCAT 1140
QY 1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGAGATCGTGTAGAAATGCTCTA 1200
DB 1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGAGATCGTGTAGAAATGCTCTA 1200
QY 1201 AATGCAACAGAGCTGCGGTTGAAGAGGTATTGTCCTGCGTGGTGGTACTGCTTTTGTG 1260
DB 1201 AATGCAACAGAGCTGCGGTTGAAGAGGTATTGTCCTGCGTGGTGGTACTGCTTTTGTG 1260
QY 1261 CCGTCCATTAAAGTCTTGTATGATATTAACCTGCTGATGATGATGATGATGATGATGAT 1320
DB 1261 CCGTCCATTAAAGTCTTGTATGATATTAACCTGCTGATGATGATGATGATGATGATGAT 1320
QY 1321 AATATATCCGCTGCTTCTTGAAGAGCTTAAAGCTCAAAATTTGCTGCAAAATGCTGCTAT 1380
DB 1321 AATATATCCGCTGCTTCTTGAAGAGCTTAAAGCTCAAAATTTGCTGCAAAATGCTGCTAT 1380
QY 1381 GAAGGTTCTATTGTTGTAGAAAGAGTTCGTGAAACCAAGATGGTTTGAATTTAATGCT 1440
DB 1381 GAAGGTTCTATTGTTGTAGAAAGAGTTCGTGAAACCAAGATGGTTTGAATTTAATGCT 1440
QY 1441 GCATCAGGAGATATGAAGACCTTAAAGCTGCTGATGATGATGATGATGATGATGATGAT 1500
DB 1441 GCATCAGGAGATATGAAGACCTTAAAGCTGCTGATGATGATGATGATGATGATGATGAT 1500
QY 1501 CGTATTGCTTACAAATGCAAGCTAGTCTTACTTCTTAACTACAGAAATGCTGCT 1560
DB 1501 CGTATTGCTTACAAATGCAAGCTAGTCTTACTTCTTAACTACAGAAATGCTGCTGCT 1560
QY 1561 ATTGCTGAAACCAAGACCTTAAAGATATGCTTATGCTGCGGTGGTATGCGGTGCT 1620
DB 1561 ATTGCTGAAACCAAGACCTTAAAGATATGCTTATGCTTATGCTGCGGTGGTATGCGGTGCT 1620
QY 1621 ATGGTGTGATGAGCGGTATGATCTAG 1647
DB 1621 ATGGTGTGATGAGCGGTATGATCTAG 1647

RESULT 2

AAT69203

ID AAT69203 standard; DNA; 4972 BP.

XX AAT69203;

XX AC AAT69203;

DT 20-AUG-1997 (first entry)

Db 1022 CTAATCGCGTAGCTATCATTAAGTACAGATGAGCGCTACAACTTCTGATTTTGATCGTG 1081
Qy 1088 AAAAATCTTCAAGACCTCTTGCAGAACTTGTGGTGGAGTAGCTGTATCCATGTTGGAG 1147
Db 1082 AAAAATTTACAAGAACCACTTGTCTAGTTAGCCGGTGTGTAGCAGTAATATAAGTTGGTG 1141
Qy 1148 CTGCTACTGAATGAATGAAGAGAGAGAGATGCTGAGAGATGCTCTAAATGCAA 1207
Db 1142 CAGCGACTGAACAGAAATTAAGAGAGATGAACCTTCGCATCGAAGATGCGTTAAATGCAA 1201
Qy 1208 CAAGAGCTCGGTGAAGAGAGATGATGCTCCGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db 1202 CCGGTGCTCAGTTGAAGAGATGATGCTTTCAGTGGAGTAGCGCTCTTGTGAACGTTA 1261
Qy 1268 TTAAGTCTCTGATGATATTAACCTGCTGATGATGATGAATGCTGCTGACTTAATATCA 1327
Db 1262 TTAAGAAAGTAGCGGCACTGAACCTTAATGCTGATGAGGA--GACTGGACGTAATATTG 1318
Qy 1328 TCGGTGCTCTCTGAAGAGCCTTTACGTCMAATGCTGCAATGCTGCAATGCTGGCTATGAAGTT 1387
Db 1319 TTTCTTGTGCTCTCGAAGAGCCTGTTGCTGCAAAATGCTTACAAATGCTGGAATGAAGTT 1378
Qy 1388 CTATTGTTGTAGAAAAGTTGCTGAACCAAAAGATGCTTTTGGATTTAATGCTGCATCAG 1447
Db 1379 CAGTTATTATTGAACCTTTAAACAGCTCTGAATTTGTACAGGATTTAATGCGGCAATG 1438
Qy 1448 GAGAATPATGAAGACCTTTAATAAGCTGGTGTCTATGATCTCTAAAGTTTACAGCTATTG 1507
Db 1439 GAGAATGGGTAGATATGTTTACCACAGGTATCATGACCCCTGTCAAAAGTAACAGCTTCTG 1498
Qy 1508 CATTACAAATGACAGTACAGTACCTCTTCTTCTTAACTACAGATGCGCTATTGCTG 1567
Db 1499 CACTTCAAAATGCGGATCTGTAGCAAGTCTTATCTTGTACTACAGAGCAGTAGTAGCAA 1558
Qy 1568 AAAAACCAGAACCTAAA 1584
Db 1559 ATAAACCTGAACAGAA 1575

RESULT 4

ID ABN71527/c
XX ABN71527 standard; DNA; 2155561 BP.
XX AC
XX ABN71527;
DT 02-JUL-2002 (first entry)
XX Streptococcus polynucleotide SEQ ID NO 10967.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus sp.
XX
XX W0200234771-A2.
PN
XX
PD 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Massignani V, Margarit Ros Yi, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR
XX

PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX Claim 8; Page 4196-4488; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Query Match 45.3%; Score 746.6; DB 24; Length 2155561;
Best Local Similarity 68.0%; Pred. No. 2.6e-161;
Matches 1072; Conservative 0; Mismatches 499; Indels 6; Gaps 2;
Qy 8 CTAAGAAATCTCTTTTGTGCTTAAAGCCCGTGAAACCTTTCCAGAGTGATAGATAAAC 67
Db 2024265 CAAAGATATTAATTTTCAGCAGATGCCGCTCAGCAATGGTGGTGTGATATTT 2024206
Qy 68 TTGCAATGCTGTAAAGTAACACTTTGGACCTTAAGCCGCTAATGCTGTTATGAAAGT 127
Db 2024205 TAGCTGATACAGTCAAAAGTAACATTAGTCTTAAAGCCGCTAATGTTCTTGAAGA 2024146
Qy 128 CTTTGGTTCGCCAGTTTATACAAAGATGCTGTATCTGTTGCAAAAGAAATGAACCTG 187
Db 2024145 CATTTGGTTCCTCTCTTAAATTAACAATGATGGTGTGACAAATGCTAAAGAAATGAGCTAG 2024086
Qy 188 AAGATAAGTTTGAATAATATGGCGCTCAAAATGGTTTAAAGAAAGTAGTCTCCAAACACTAGCG 247
Db 2024085 AAGATCACTTTGAATAATATGGAGCTTAACTTGTGTGAGAAGTGGCTTCAAAACACTAATG 2024026
Qy 248 ATATTGCTGTGATGGAACCTACACAGCAACAGTCTTGTGACACAGCTATTATCTGTAAG 307
Db 2024025 ATATTGAGGGGATGGCACTACAACTGCTACTGTTTGTGACCCCAAGCTATTGTACGGGAAG 2023966
Qy 308 GTGTAAACTTTGAGCAGCTGGTGTGTAATCTTATGCGCATTAAACGTTGGCATAGATAAG 367
Db 2023965 GTCTTAAATAATGTAACCTGCAGGGGCAAAATCCGATTGGCATTGCTGCTGGTATTGGAACAG 2023906
Qy 368 CTGTTGTTGCTGTATTCTAAGAACTAAGCGACATTAACAAAGCCCTACTCTGTACACAAAG 427
Db 2023905 CTGTTTACGACAGCTTTGAAGAGCTTAAAGAGATTGACACACCACTTTTCAGGCAAGAG 2023846
Qy 428 AAATAGCTCAAGTTGGAAACCATTTCTGCAAACTCTGTATACAACTAGTAATATCATAG 487
Db 2023845 CTATTGCTCAAGTTGCGGCTGTGTCTTTCAGCTTCTGTA--AAAAGTTGGGGAATACATTT 2023789
Qy 488 CTGNAGCTATGGCTAAAGTTGGAAAGAGGAGTGTATACAGTTTGAGAGAGCTAAAGGTC 547
Db 2023788 CTGAAGCTATGGAGCGGTGGTGAATGATGGTGTATACATTATGAAGATCGGAGGTA 2023729
Qy 548 TTGAACCTACATTAGATGCTGTTCAAGGAATGAAGTTTGAACCGTGGCTCTCTCTCCAT 607
Db 2023728 TGGAAACAGAGCTTGAAGTTGTGAAGGAATGCASTTTGACCGTGGTACTTGTACAGT 2023669
Qy 608 ACTTTGTAACCTAATCCCTGAGAAATGTTGTGAACCTGTGATAACCCCTTATATCTTTGTA 667
Db 2023668 ATATGGTAACCTGATAACGAGAAATGGTCTCTGAACTTGAGATCCGATATCTCTTATTA 2023609

QY	668	ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTTAGAACAAAGTTGCTAAAG	727
Db	2023608	CAGATAAGAAAATTTCAAATATCCAGAAAATTTTACCAITATTAGAGAGAGTTCTTAAAA	20233549
QY	728	TAAACCGTCCACTCCTTTATTATTGCTCAAGACGTAGAGGTGAAGCACTTGGCAACTTGG	787
Db	2023548	CAAAATCGTCGTGCTTAATCATCGCTGATGATGTTGATGAGAGAGCTCTCCCAACGCTTG	20233489
QY	788	TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAAGCTCTCTGTTTTGGTG	847
Db	2023488	TTCTTAAACAAAATTCGTGGAACTTCAATGTGCTAGCTGTAAAGCGCTGGATTTGGTG	20233429
QY	848	AACCCCGTAAAGCTATGCTTGAAGATTTGCTATCCTTACTGAGAGAGAGCAATATTTG	907
Db	2023428	ATCGTCGTAAGCCATGCTGGAAGATATGCTATPCTTAACAGAGAGAACTGTGCTTACTG	20233369
QY	908	AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTTTAGGAAACAGCTTAAACGCTG	967
Db	2023368	AAGACCTTGGTTAGACTTAAAGATGCTACTATGCAAGTTTTAGCACAGTCTGCTAAAG	20233309
QY	968	TAGTTATGACAAAGAAAATCTACTATCGTTGATGGTGTGGAATAATCAGAAGATATTA	1027
Db	2023308	TAAAGTAGATAAAAGATTTCTACTGTATTGTCCAAGGTGCGGTGACTCATCAGCAATTG	2023249
QY	1028	AAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAAGCTCAGATTAATGATCGTG	1087
Db	2023248	CTAATCGCTAGCTATCATTTAAGTCAAGATGGAGGCTACACTTCTGATTTTGTATCGTG	2023189
QY	1088	AAAACTTCAAGAACGCTTTGCCAAAATCTTGTGTGAGTAGCTGTATTCCATGTTGGAG	1147
Db	2023188	AAAAATACAAAGACGACTTGCTAAGTTAGCCGGTGGTGTAGCAGTAATTTAAGTTGGTG	2023129
QY	1148	CTGCTACTGAACTGAATCAAGAGAGAGAGAGATCGTGTAGAGATGCTCTAAATGCAAA	1207
Db	2023128	CAGCGACTGAACAGAAATTAAGAGAGATGAACCTTCGCATCGAAGATGCGTTAAAAATGCAA	2023069
QY	1208	CAAGAGTCGGGTTGAAGAAGGTATTGTCCCTCGTGGTGTACTGCTTTTGTCCGCTCCA	1267
Db	2023068	CGCGTCTGCAGTTGAAGAAGTATTGTTTCAGTGGAGGTACGGCTCTTTGTGAACGTTA	2023009
QY	1268	TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACCTTGTGCACTTAATATCA	1327
Db	2023008	TTGAAAAGATGAGCGCACTGAAACTTAATGGTGATGAGGA--GACTGCAAGCTAATATG	2022952
QY	1328	TCCGTGCTTCTCTTTGAAGAGCCCTTACGTCAAATTCCTGCAAAATGCTGCTATGAAGTT	1387
Db	2022951	TTCTTGTGCTCTCGAAGAGCCGTGTCGCAAAATGCTTACAAATGCTTACAAATGCTGATGAAGTT	2022892
QY	1388	CTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG	1447
Db	2022891	CAGTTATTAATTGAACGTTTTAAACAGCTCTGAATTTGTTACAGGATTTAATGCGGCAATG	2022832
QY	1448	GAGAAATGAAGACCTTATTAAAGCTGGTGTCTATTGATCCTTAAAAAGTTACAGTATTG	1507
Db	2022831	GAGAAATGGGTAGATATGGTTACACACAGGTATCATTTGACCTCTCAAGAGTAAACGCTTCTG	2022772
QY	1508	CAATTACAAAATGACAGCATAGTAGCCTCTTACTTCTTAAGTACAGAAATGCGCTATTGCTG	1567
Db	2022771	CACCTTCAAAATGGGCATCTGTAGCAGAGCTTTATCTTGACTACAGAGCATGTAGTACAA	2022712
QY	1568	AAAAACAGAACCTAAA	1584
Db	2022711	ATAAACCTGAACAGAA	2022695

Qy	308	GTGTAACCTTTAGCAGCTGGTTCGTAATCCTATGCCATTAAACGTGGCATAGATAAAG	367
Db	49580	GCATGAAAAACGGTTGTCGAGCATGAATCAAATGGATCTAAAGCTGGTATTGATAAAG	49639
Qy	368	CTGTGTTGCTCTGTACTATAAGAACTAAAGCGACATTAAAGAGCTACTCGTCAGCAAAAAG	427
Db	49640	CGGTACGGCGCGCTTTTGAAGAAATCGTGTATCTTCGACACTCTTAATGACATTAAG	49699

RESULT 5
AAF28548
ID AAF28548 standard; DNA; 96109 BP.
XX
XX AAF28548;
XX
DT 04-APR-2001 (first entry)

QY	428	AAATFAGCTCAAGTTGGAACCAATTTCTGCAACTCTTGATACAAATAGGTAAATATCATAG	487
Db	49700	CGATTGCTCAAGTTGGCTCAATTTACAGAACTCAGATGCTACCAITTTGGTGAAGTTATCT	49759
QY	488	CTGAAGCTATGGCTAAAGTTGGAAAAAGGAGGTGTTATCACAGTTGAGGAGCTAAAGGTC	547
Db	49760	CCAAGGCAATGGAAACAGTTGGCAAAACAGGCGTTATCACAGTTGAAGAGGGTTTCAGGTT	49819
QY	548	TTGAAACTACATFAGATGCTGTTGAGGAGTGAAGTTTGACCGTGGCTACCTCTCTCCAT	607
Db	49820	TTGAAGATGCGTTGAAAGTAGTTGAGGCGATGAGTTTGACCGTGGCTATATCAGCCCTT	49879
QY	608	ACTTTGTAACTAATTCCTCGAGAAAATGTTTGTGAACTTGATAACCCCTTATATCTTTGTA	667
Db	49880	ACTTTGCCAATAACAGACAGCTTGACTTGTGAAATTTGACATCCATTTATCTTTTGG	49939
QY	668	ATGAGAAAAAGATTCTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTTGCTAAAG	727
Db	49940	TCGATAAAAAAATCTCAAAATATTCGTGAGATTTGTGCCACTACTTGAAAAAAGTGATGAAA	49999
QY	728	TAAACCTGCCACTCTTATTTATTTGCTCAAGACGTAGAAGCTGAAGCACTTGCACACTTG	787
Db	50000	CCAGCCGTCCGTATTAATCATTCGTCAAGATGTTGAAAATGAAGCATTTGGCAACATTGG	50059
QY	788	TAGTCAATAAGCTCCGTGGAGCACTCAAGATTTGTAGCGCTAAAAGCTCTCTGTTTGGTG	847
Db	50060	TTGTCAATACTTTGCGTGGCGGATTAATAACTTTGTGCGGTTAAAGCCCAAGTTTGGTG	50119
QY	848	AAGCCGTGAAGCTATGCTTGAAGATATTGCTATCCTTACTGAGAGAGAAAGCAATATTG	907
Db	50120	ATGCCCGTAAAGCCATGCTTCAAGACATTTGCAATTTTAAACAGTGGCGTGTATCTCAG	50179
QY	908	AAGATCGTGGTATAAAGCTTGAAATCTAAGCTTGTCTTTTAGGAACAGCTAAACGTG	967
Db	50180	AAGAAGTGGGCTTAAAGCCTTGAGACTGCTGAGATTTGAGCAATTTGGGTACTGCCAAAAAG	50239
QY	968	TAGTTATTGACAAAAGAAATACTACTATCGTTGATGCTGTGGAATAATCAGAAGATTTA	1027
Db	50240	TAACTATTGTAAGAAACACCCGTCATTTGATGCTGAGGTGATAAGCCAGCATTG	50299
QY	1028	AAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGATCGT	1087
Db	50300	AGGCGGTGTAGATCAATTCGTCGTCAGGTTGAAGAAATCAATCTCTGACTACGACAAAG	50359
QY	1088	AAAACCTTCAGAACGCTTGCAAACTTTGTTGTGAGTAGCTGTTTATCCATGTTGGAG	1147
Db	50360	AAAAAATTCAGAGCGTGTGCAAAACTATCAGGCGGTGTGTCAGTCAATCAAGGTGGGTG	50419
QY	1148	CTGCTACTGAACTGAAATGAAGAGAAAGAGTCTGTAGAGATGCTCTAAATGCAA	1207
Db	50420	CGGCACTGAAACTGAGATGAAGAGAAAGAACCGGTGAGACGATGCATTCACGCTA	50479
QY	1208	CAAGAGCTCGGTTGAAGAAAGTATTGTCCTGTGTGTGTGCTGTTTGTGCGTCCCA	1267
Db	50480	CCCGTGTGCTGTTGAAGAGGCTGTGTGCTGTGTGTGTGTGCTGCTGCTGCTGCT	50539
QY	1268	TAAAGTCTTGATGATATTAACCTGCTGATGATGATGACTTGTGAGCTTAATATCA	1327
Db	50540	TATCTGCAATTAAGCGACCTTAAGAGGTGATATGAAGACCAAAATGACGAGCATCAATTC	50599
QY	1328	TCCGTGCTTCTTTGAAGAGCCTTTACGTCAAAATGCTGCAAAATGCTGCTATGAAGGTT	1387
Db	50600	TACGCGTGCAGTGAAGACACACTTCCGCAAAATCGTCAGTAAACGAGGTGATGAGGCTT	50659
QY	1388	CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGTTTGTGATTTAATGCTGCATCAG	1447
Db	50660	CTGTAAATTGTTAATCAAGTTAAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	50719
QY	1448	GAGAAATGAAGACCTTATTAAGCTGTGTGTCATGCTCTTAAAAAAAGTTACAGGTATTG	1507
Db	50720	GTGAGTATGGCGATGCTTGAAATGGGTATTTTAGATCCCTGCCAAAGTAACCCGTTTCAG	50779

QY	1508	CATTACAAAATGCAGCATCAGTAGCTCTTACTTCTTAACCTACAGAAATGCCCTATTGCTG	1567
Db	50780	CATTAGAACATGCTGCTTCTGTCGAGGTTTGTGATGTTGACCACTGAGGTGAATATCCG	50839
QY	1568	AAAAACACAGAACCTAAAAAAGATATGCCCTATGCCCTGCGGTGGTATGGTGGTATGGGTG	1627
Db	50840	ATAAACACAGCACCT-----AGAAGCACAATGCCAGAGGTGGTATGGTGGTATGGGTG	50893
QY	1628	STATGACCGGTATGTACTA 1646	
Db	50894	GTATGCGCGGTATGATGTA 50912	

RESULT 6
ABQ70705
ID ABQ70705 standard; DNA; 1901 BP.
XX ABQ70705;
AC ABQ70705;
XX 29-AUG-2002 (first entry)
XX Listeria monocytogenes 4b contig DNA sequence #647.
XX Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX Listeria monocytogenes 4b.
XX WO200228891-A2.
XX 11-APR-2002.
XX 04-OCT-2001; 2001WO-FR03061.
XX 04-OCT-2000; 2000FR-0012697.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from Listeria species, useful for detection, treatment and prevention of infection, also related polypeptides, antibodies and modulators -
XX Claim 14; SEQ ID 3518; 180pp; French.
XX The present invention relates to nucleic acid sequences (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes and primers for identification and/or detection of Listeria (e.g. as contaminants in foods, or mutational analysis) and for analysis of gene expression. Proteins encoded by the nucleic acid sequences can be used to screen for compounds that modulate gene expression, replication and pathogenicity of Listeria (potential therapeutic agents), also for treating infections by Listeria, and are useful as immunogens in anti-Listeria vaccines.
XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1901 BP; 660 A; 345 C; 435 G; 461 T; 0 other;
SQ

Query Match 44.5%; Score 732.2; DB 24; Length 1901;
Best Local Similarity 66.5%; Pred. No. 6.1e-159;
Matches 1081; Conservative 0; Mismatches 538; Indels 6; Gaps 2;
QY 8 CTAAGAAATCTCTTTTGATGCTAAAGCCGCGGAAAACTTTTACAGAGGTGATAGATAAC 67
Db 232 CAAAGATATTAATTTAGTAGATGCTCGTCGCGCATGTTAGTGTGTCGACCAAT 291
QY 68 TTGCAATGCTGTAAAGTAAACACTTGGACCTTAAAGCGCGGTAAATGTCGTTATTGAAAGT 127

Db 292 TAGCAACGCGTAAAGTAACACTTGGCCAAAAGTCGTAATGTTGTTTAAAGAAA 351
QY 128 CTTTGGTTCACAGTATTAACAAGATCGTGTATCTGTGCAAAAAGAAATGAACCTG 187
Db 352 AATTGGTTCCTCCGTTAAATTAACAATGATGGGTAACGATTGCAAAAAGAAATCGAATTAG 411
QY 188 AAGATAAGTTTGAATAATGGGCGCTCAAAATGGTTAAAGAAAGTAGCTCCCAAACTAGCG 247
Db 412 AAGACCCATTTGAAAATATGGGAGCAAACTTGTATCTGAAGTTGCTTCTAAACCAATG 471
QY 248 ATATTGCTGGTGATGAACTACAACAGCAACAGTCTCTGCAACAGCTATTTATTCGTGAAG 307
Db 472 ATGTTGCTGGGACGCAACTACAACAGCTACCGTTTTAGCGCAAGCAATGATTTCAAGAAG 531
QY 308 GTGTAAACTTTAGCAGCTGGTCTGAATCCTATGSCCAATTAACGTTGGCATAGATAAG 367
Db 532 GCTTGAACCACTAAGCTGAGCAAAATCCAGTAGGCGTTGCGCCGGTATCGAAAAAG 591
QY 368 CTGTTGTTGCTGTACTAAAGAACTAAGCGCAATTAACAAGCCTACTCGTGACCAAAAAG 427
Db 592 CGGTAGCAACACTATCGAGAAATTAAGAGCTATCTTAACCAATTAAGCAAGAGT 651
QY 428 AATAGCTCAAGTTGAAACATTTCTGCAAACTCTGATACAACAATAGTAATATCATAG 487
Db 652 CTATCGCTCAAGTTGCTGCTATTTCTTC---TGGTGATGAAGAAGTTGGTAAATTAATCG 708
QY 488 CTGAGCTATGCTAAAGTTGAAAGAGAGGTGTTATCACAGTTAGGAGCTAAAGTTC 547
Db 709 CAGAAGCAATGAACGTGTGTGAACGCGGGTTATCACTATTGAAGAAATCCAAAGGCT 768
QY 548 TTGAACACTACATAGATGTTGTTGAAGGAATGAAGTTTGACCGTGGCTACTCTCTCCAT 607
Db 769 TTGCAACAGAAATAGATGTAGTAGAGGTATGCAATTTGACCGTGGCTACATAGCCCAT 828
QY 608 ACTTTGTAATACTCTGAGAAAATGGTTTGTGAACCTTCATAACCCCTTATATCCTTTGTA 667
Db 829 ACATGTAACCTGATTCGACAAAATGGAAGCAGTCTCTGAAAAAACCATATATTTTAATTA 888
QY 668 ATGAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAGTTGCTAAG 727
Db 889 CAGACAAAATAATCAACACATCCCAAGAAATCTTACCAGTTTTTAGAACAAGTTGTTCAAC 948
QY 728 TAAACCGTCCACTCTCTTATTATGCTGAAGAGCTAGAAGTGAAGCACTTGAACACTTG 787
Db 949 AAGTGTCTCCATGTTAATCATTTGCGAGATGTTGAAGGGAGAGCTCAAGCAACTCTTG 1008
QY 788 TAGTCAATAAGCTCCGTGAGCACTCCAAAGTTGTAGCCGTAAGAACTCTGTTTGGTG 847
Db 1009 TACTAAACAACTTCGCGGAACATTCACGTAGTAGCCGTGAAGAGCTCTGTTTGGTG 1068
QY 848 AAGCGCGTAAAGCTATGTTGAAGATATGCTATCTTACTGAGGAGAGCAATATTG 907
Db 1069 ATGCTGTAAGCAATGCTAGAGATATTGCTATTTTAAACGCGGAGCAAGTCAACAG 1128
QY 908 AAGATCGTGTATAAGCTTGAAGATGAAGCTTGTCTTTTAGCAACAGCTAAACGTG 967
Db 1129 AAGACCTAGGCTTAGAATTAAGAAACAGCAACAGTTGATCACTTGAACAGCGCAACAG 1188
QY 968 TAGTTATTGACAAAGAAATACTATCTATGTTGATGGTCTGGAATAATCAGAAAGATATTA 1027
Db 1189 TAGTTGTAACAAAAGATGATACAACATCGTAGAAGGAGCAGGCGATTCCACACAAATTA 1248
QY 1028 AAGCTCGAGTTAAACAAATCTGTGACAAATTAAGAAACAGCTCAGATTATGATCGTG 1087
Db 1249 GTGCTCGCTAAACCAATCCGTCGCGCAATGGAAGAACTACTTCTGAATTTGATAGAG 1308
QY 1088 AAAAACTTCAAGACGCTTTGCAAACTTTGTTGGTGGAGTAGCTGTTATCCATGTTGAG 1147
Db 1309 AAAAAATTACAAGAACGTTTAGCAAACTTGCAGGTGGGTAGCTGTTGTCAAGTCGCG 1368
QY 1148 CTGCTACTGAACTGAATGAAGAGAGAGAGATCGTGTAGAGATGCTTAATATGCA 1207

Db 1369 CTGCAACTGAAACAGAGCTAAAGAAAGCTAAATTCAGTATTGAAGATGCGCTTAACCTTA 1428
QY 1208 CAGAGCTCGGTTGAAGAGGTATTTGTCCTGGTGGTGGTACTGCTTTTTCCTCCCTCCA 1267
Db 1429 CTCGCGCAGCTGTAGAAGAGGTATCTAGCTGGTGGTGGTACTGCTCTTTGTAAGTATTT 1488
QY 1268 TTAAAGTCTCTTGATGATATTAAACCTGCTGATGATGAACCTTGTGGACTTAAATATCA 1327
Db 1489 ACATTAAGTAGCAGACACTAGAA---GCAGAGGTGACGTAGAACTGGTATCAACATCG 1545
QY 1328 TCCGTCGTTCTCTTTGAAGAGCCTTTACGTCAAATTCCTGCAATGCTGGCTATGAAGGTT 1387
Db 1546 TCGTTCGTTCCCTTAGAAGAACAGTTTCGCGCAATCGGCATTAACGCTGACTTTGAAGTT 1605
QY 1388 CTATTCTTTAGAAAAAGTTCTGTAACCAAAAGATGTTTTGGAATTAATGCTGCATCAG 1447
Db 1606 CGGTTATCGTTGAACGCTTTGAACACGAAAGCAGTTGGCGTTGGTTTCAACGCGACTAATG 1665
QY 1448 GAGAATATCAAGACCTTTATTAAGCTGGTGTCTATGATCTTAAAAAGTTTACACGTATTG 1507
Db 1666 GCGATGGGTAAACATGATTGACGCTGGTATTGCGATCCAAAGAAAGTAAACAGCTTCG 1725
QY 1508 CATTAACAAATGCGATCAGTAGCCTCTTACTTCTAACTACAGAAATGCGCTATTGCTG 1567
Db 1726 CACTTCAAAACGATCATCTGTTGCTGCACTTCTATTAACTACAGAAAGCAGTCGTAGCAG 1785
QY 1568 AAAACCCAGAACTTAAAAAGATATGCTTATGCTGGCGGTGGTATGGTGGTATGGTG 1627
Db 1786 ACAACCCAGACGAAACCGGCCAGAGCTGTTCTCTGATATGGAATGGGTGGCATGGCG 1845
QY 1628 GTATG 1632
Db 1846 GCATG 1850

RESULT 7

ABQ67193

ID ABQ67193 standard; DNA; 213251 BP.

AC ABQ67193;

DT 29-AUG-2002 (first entry)

XX Listeria innocua contig DNA sequence #6.

XX Antibacterial; Listeria; food contamination; mutational analysis;

XX infection; ds.

XX Listeria innocua.

XX WO200228891-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-FR03061.

XX 04-OCT-2000; 2000FR-0012697.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Kunst F, Glaser P;

XX WPI; 2002-332479/37.

XX New genomic sequences from Listeria species, useful for detection, treatment and prevention of infection, also related polypeptides, antibodies and modulators

XX Claim 5; SEQ ID 6; 180pp; French.

XX The present invention relates to nucleic acid sequences

CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes

CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 213251 BP; 70939 A; 36026 C; 43257 G; 63029 T; 0 other;

Query Match 44.3%; Score 729; DB 24; Length 213251;
Best Local Similarity 66.4%; Pred. No. 1.5e-157;
Matches 1079; Conservative 0; Mismatches 540; Indels 6; Gaps 2;

QY 8 CTAAGAGAAATCCTTTTGTGCTAAAGCCGCTGAAAGAACTTTCACGAGGTGTAGATAAAC 67
DB 135668 CAAAGATATTAATTTAGTGAAGATGCTGCTGCGCATGTTACGTGGTGTGACCAAC 135727

QY 68 TTGCAAAATGCTGTAAAGTAACACTTGGACCTAAAGCCGCTTAATGCTGTTATTGAAAAGT 127
DB 135728 TAGCAACGCGTAAAGTAAGCTTGGCCCAAGAGTCTGATGTTGTTTGTAGAAAAA 135787

QY 128 CTTTGGTTCGCCAGTATTACAAAAGATGGTGTATCTGTGCAAGAAATTTGAACCTTG 187
DB 135788 AATTCGGTTCTCCGTTAATTAACAAATGATGGTGAACGATTTGCAAGAAATCGAATTAG 135847

QY 188 AAGATAAGTTTGAAGATGCGGCTCAAATGGTTAAAGAGTAGTCCCAAACTAGCG 247
DB 135848 AAGACCATTTGAACAATGGGCGCAAACTTGTATCTGAAGTTGCTTCTAAACCAATG 135907

QY 248 ATATTGCTGGTGAAGTAACACTACAAAGCAAGCTTTCGACAAAGCTATTATTTCGTGAAG 307
DB 135908 ATGTTGCTGGGAGCAACTAGCACTGCTACCGTTTTAGCACAGCAATGATTCAAGAAG 135967

QY 308 GTGTAAACTTTAGAGCTGGTGTGTAATCTGATGGCCATTAACGTTGGCATAGATAAG 367
DB 135968 GCTTGAAGAAACGTAACAGCTGGAGCAATCCAGTAGCGTTTCGTGGTATCGAAAAAG 136027

QY 368 CTGTTGTTGCTGTACTAAAGAACTAAGAGCACTTACAAAGCTTACTCGTGACCAAAAG 427
DB 136028 CCGTAGCACTGCTATTGAAGAAATTAAGCTATCTTAACCAATTTGAAGCAAGAGT 136087

QY 428 AAATAGCTCAAGTTTGAACCAATTTCTGCAACTCTGATACAACTAGGTAATATCATAG 487
DB 136088 CTATCGCTCAAGTTTGTGCTCAATTTCTTC--TGGTGATGAAGAGTTGGTAAATTAATCG 136144

QY 488 CTGAGCTATGGCTAAAGTTGGAAGAGAGTGTGTTTACAGTTGAGGAGCTAAAGTTC 547
DB 136145 CAGAAGCAATGGAAACGTTTGGTAACGATGGTGTATTACTATTGAAGAAATCCAAAGGCT 136204

QY 548 TTGAAACTACATTAAGATGCTGTTCAAGGAATGAAGTTTGACGGTGGCTACCTCTCCAT 607
DB 136205 TTGCAACAGAAATAGACGTAGAGGATGCAATTTGACCGTGGCTACACTAGTCTCTT 136264

QY 608 ACTTTGTAACCTTAACCTGAGAAAAATGGTTTGTGAACCTTGAACCTTATATCTCTTTGTA 667
DB 136265 ACATGGTAACCTGATTCGCACAAATGGAAGCAGTCTTGAAGAACCATACATATTTAATTA 136324

QY 668 ATGAGAAAAAGATTAAGCATGAAAGCATGCTACCAATCTTAGAACAGTTCGTAAG 727
DB 136325 CAGATAAAAAAATCAACAACTCCAAAGAACTTATACAGTTTGAACAGTGTGTTCAAC 136384

QY 728 TAAACCGTCCACTCTTATTATTCGTAAGAGTGAAGGTGAAGCACTTCCAACTTG 787
DB 136385 AAGTCTGCTCAATGTTAATATCGGGAAGATGTTGAAGGGAAGCTCAACCACTCTAG 136444

QY 788 TAGTCAATAGCTCCGCGGAGCACTCCAAAGTTGTAGCGGTAAAAAGCTCCTGTTTGGTG 847
DB 136445 TACTAAACAAACTTCGCGGAACATTTAAGCTAGTTGCAAGTCAAGCTCTGTTTCGGTG 136504

QY 848 AACGCGTAAAGCTATGCTTGAAGATATGCTATCTTACTCGAGGAGAAAGCAATATTG 907
DB 136505 ACCGCGTAAAGCAATGCTAGAGATATGCTGTTTAACTGGTGGCAAGTATACAG 136564

QY 908 AAGATCGTGTATTAAGCTTTGAAAAATGTAAGCTTGTCTTTTGAAGCAAGCTAAAGCTG 967
DB 136565 AAGACCTAGGTTTAGAACTGAAAAACAGCAACAGTGTATCAACTTGGAAACAGCTAAACAAG 136624

QY 968 TAGTTATTGAAAGAAATTAAGTAACTACTATCGTTGATGGTCTGGAATCAGAGATATTA 1027
DB 136625 TAGTTGTAACAAAAGATGATACAACTATCGTAAGAGAGCAGCGATTCACACAAATCA 136684

QY 1028 AAGCTCAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAGCTCAGATTATGATCGTG 1087
DB 136685 GCCTCGCTAAACCAATCCGTCGCAATGGAAGAACTACTTCTGAATTTGATAGAG 136744

QY 1088 AAAAACTTCAAGAACGCTTTGCAAACTTGTGGTGGAGTAGCTGTTTATCCATGTTGGAG 1147
DB 136745 AAAAAATTACAAGAACGTTTAGCAAACTTCGACGTGGAGTAGCTGTTGTTTAAAGTCGCG 136804

QY 1148 CTGCTACTGAACTGAAATGAAGAGAGAGAGATCGTGTAGAGATGCTCTAAATGCA 1207
DB 136805 CTGCAACTGAACAGAGCTAAAGAAACGTAATTTACGTATTGAAGATGCCCCCTTAACTCTA 136864

QY 1208 CAAGAGCTGCGGTTGAAGAGGTTATTGCTCCTGCTGGTGGTACTGCTGTTTGTCCGCTCCA 1267
DB 136865 CTGTCGACAGTAGAAGAGAGTATCGTAGCTGGTGGTGGTACTGCTGCTGTTGTAAGTATTT 136924

QY 1268 TTAAGTCTTGTGATGATTTAAACCTGCTGATGATGATGAACCTTGTGGACTTAAATATCA 1327
DB 136925 ACAATAAGTAGCAGCACTAGAA---GCAGAAGCGGATGTAGAAACAGGTATCAACATCG 136981

QY 1328 TCGTGTCTCTCTTGAAGAGCCTTTAGCTCAAAATGCTGCAATGCTGGCTATCAAGGTT 1387
DB 136982 TGCTTCGTTCCCTAGAGAAACCAAGTTCGTGAGATCGCGCATACGCTGGAGCTTGAAGGTT 137041

QY 1388 CTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGTTTGGATTTTAAATGCTGCATCAG 1447
DB 137042 CTGTTATCGTAGAACGCTTGAACACAGAGAGTGGCGTTGGTTTCAACGACCTAATG 137101

QY 1448 GAGAAATGAGACCTTTATTAAAGCTGGTGTGATGATTCCTTAAAGAAAGTTACAGTATG 1507
DB 137102 GCGAATCGGTAAACATGATTGAGCTGTTGATGATTCGATCAACAAAAAGTTACTCGTCCG 137161

QY 1508 CATTAACAAATGAGCATCAGTAGCCTCTTACTTCTAACTACAGATGCGCTATGCTG 1567
DB 137162 CATTAACAAACGCTTCACTCTGTTGCTGCTCTCTTATTAACTACCGAAGCTGTTGAGCAG 137221

QY 1568 AAAAACCCAGAACCTTAAAAAGATATGCTATGCTGCGGTGGTATGGGTGATGGGTG 1627
DB 137222 ATCAACAGATGAACCGGCCAGCAGCCAGATATGGGAATGGCGGATGGCG 137281

QY 1628 GTATG 1632
DB 137282 GTATG 137286

RESULT 8
ABQ69245/C
ID ABQ69245 standard; DNA; 3011208 BP.
XX
AC ABQ69245;
XX
DT 29-AUG-2002 (first entry)
XX
DE *Listeria innocua* DNA sequence #684.
XX
KW Antibacterial; *Listeria*; food contamination; mutational analysis;
XX infection; ds.
XX
OS *Listeria innocua*.
XX
PN W0200228891-A2.

XX PD 11-APR-2002.
XX PF 04-OCT-2001; 2001WO-FR03061.
XX PR 04-OCT-2000; 2000PR-0012697.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from *Listeria* species, useful for detection, PT
treatment and prevention of infection, also related polypeptides, PT
antibodies and modulators -
XX Claim 5; SEQ ID 2058; 180pp; French.
XX The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: the sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;
Query Match 44.3%; Score 729; DB 24; Length 3011208;
Best Local Similarity 66.4%; Pred. No. 3.3e-157;
Matches 1079; Conservative 0; Mismatches 540; Indels 6; Gaps 2;
QY 8 CTAAAGAAATCTTTTGGATGCTAAAGCCGCGGAAAGAACTTTCCAGGCTGTAGATAAC 67
DB 2196334 CAAGAATATTAATTTAGTAGAATGCTCGTCGCGCAATGTTACGTGTGCGCAAC 2196275
QY 68 TTGCAATGCTGTTAAAGTAACTTGGACCTAAAGCCGCTAATGTCGTTATTGAAAAGT 127
DB 2196274 TAGCAACGCGAGTAAAGTAAAGCTTGGCCCAAGAGTGTATGTTGTTTGAAGAAA 2196215
QY 128 CTTTGGTCCCGAGTTATCAAAAGATGGTGTATCTGTTGCAAAAGAAATGAACTTG 187
DB 2196214 AATTCGGTCTCCGTTAATTAATAAATGATGGTGTAAACGATTGCAAAAGAAATCGAATTAG 2196155
QY 188 AAGATAAGTTTGAATAATGGCGCTCAATGCTTAAAGAACTAGTCCCAAACTAGCG 247
DB 2196154 AAGACCAATTTGAACAATGGCGCAAACTGTATCTGAAGTGTCTTCAAAACCAATG 2196095
QY 248 ATATTGCTGGTGATGGAACCTCAACAGCAACAGCTCCTTGCACAAGCTATTATCTGTAAG 307
DB 2196094 ATGTTGCTGGGACGGAACCTCAACTGCTACCGTTTACCAAGCAATGATTCAGAG 2196035
QY 308 GTGTAAACTTTAGCAGCTGTGCTAATCTTATGCGCAATTAACGTTGGCAATAGTAAG 367
DB 2196034 GCTTGAAAACCGTAAACGCTGGAGCAAAATCCAGTAGGGCTTCGTCGTGGTATCGAAAAG 2195975
QY 368 CTGTTGTTGCTCTTACTAAGAACTAAGCGACATTAACAAGCTTACTCTGTGACCAAAAAG 427
DB 2195974 CCGTAGCACTCTATTAGAAATTAAGAGCTATCTTAACCAATTTGAAGCAAAAGCT 2195915
QY 428 AAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAAATAGTAGTAATATCATAG 487
DB 2195914 CTATCGCTCAAGTTGCTGCAATTTCTTCTC---TGGTATGAAGAACTTGGTAAATTAATCG 2195858
QY 488 CTGAAGCTATGCTAAGTTGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGTTC 547

DB 2195857 CAGAAGCAATGGAACGTTGGTAAACGATGGTGTATTACTATTGAAGAAATCCAAAGCT 2195798
QY 548 TTGAAACTACATTAGATGGTGTGAAGGAATGAAGTTTCACCGTGGCTACCTCTCTCCAT 607
DB 2195797 TTGCAACAGAAATTAGACGTAGTAGAAGGTATGCAATTTGACCGTGGCTACCTAGTCCCT 2195738
QY 608 ACTTTGTAACTAATCCTGAGAAAAATGGTTTGTGAACCTTGATTAACCCCTTATATCTTTGTA 667
DB 2195737 ACATGGTAATGATTCGACAAAATGGAAGCAGTCTCTGAAAACCATACATATTTAATTA 2195678
QY 668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACCAAGTTGCTAAAG 727
DB 2195677 CAGATAAAAAAATCAACACATCCAGAAAATCTTATCAGTTTATAGAACAAAGTTGTCAAC 2195618
QY 728 TAAACGCTCCACTCTTATTATTGCTGAAGCTGAGAGGTGAAGCTGAGCACTTGGCAACTTG 787
DB 2195617 AAGTGTCTCAATGTTAATTAATGCGGGAAGATGTTGAAGGGGAAGCTCAAGCAACTCTAG 2195558
QY 788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCCTAAAAGCTCTCTGGTTTGGTG 847
DB 2195557 TACTAAACAAACTTCGCGGAACATTTAAGTAGTTGCACTGAAAGCTCTCTGGTTTCGGTG 2195498
QY 848 AACCGCTAAAGCTATGCTTTGAAGATATTGCTATCTTACTGAGAGAGCAATATTG 907
DB 2195497 ACCGCGCTAAAGCAATGCTAGAGATATTGCTGTTTAACTGCTGACAAAGTAATCACAG 2195438
QY 908 AAGATCGTGTATAAGCTTTGAAATGTAAGCTTGTCTTTCTTTAGCAACGCTAAACGCTG 967
DB 2195437 AAGACTGTGTTTAGAAGCTGAAAACAGCAACAGTTGATCAACTTGGAAACGCTAACAAAG 2195378
QY 968 TAGTTATTGACAAAAGAAATATCTATCTGTTGATGGTCTCGAAAATCAGAAATATGATTA 1027
DB 2195377 TAGTTGTAACAAAAGATGATACAACTGCTAGAGGACGAGCATTCACACAATCA 2195318
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGACTCAGATTATGATCGTG 1087
DB 2195317 GCGCTCGCTAAACCAAAATCCGTGCGCAATGCGCAAAATGCAAGAACTACTCTTGAATTTGATAG 2195258
QY 1088 AAAAACTTCAAGAACGCTCTTGCAAACTGTTGTTGGAGTAGCTGTATCCATGTTGGAG 1147
DB 2195257 AAAAATTACAGAACGTTTAGCAAACTTGCAAGGTGGAGTAGCTGTGTTTAAAGTCGGG 2195198
QY 1148 CTGCTACTGAAACTGAAATGAAAGAGAGAGGATCGTGTAGAAGATGCTCTAAATGCAAA 1207
DB 2195197 CTGCAACTGAACAGAGCTTAAAGAACGTAATTAAGTATTGAAGATGCTTAACTCTA 2195138
QY 1208 CAAGAGCTGGGTTGAAGAGTATTGCTCCCTGGTGGTGGTACTGCTTTTGGCCGCTCCA 1267
DB 2195137 CTCGTGCAAGCAGTAGAAGAGGTATCGTAGCTGGTGGTGGTACTGCTTGTAAAGTATTT 2195078
QY 1268 TTAAGTCTTGTATGATTAATAACCTGCTGATGATGATGAACTTCTGCTGCACTTAATCA 1327
DB 2195077 ACATTAAGTAGCAGCACTAGAA---GCAGAAAGCGATGTAGAAACAGGTATCAACATCG 2195021
QY 1328 TCCGCTGTTCTCTTGAAGAGCCCTTTACGTCATAATGCTGCAATGCTGCTATGAAAGTT 1387
DB 2195020 TGCTTCGTTCCCTAGAGAACCCAGTTCCGTAGATCGCATAAACGCTGCACTTGAAGTT 2194961
QY 1388 CTAATTGTTGTAAGAAAAGTTGCTGAACCAAAAGATGGTTTGGATTTAATGCTGCATCAG 1447
DB 2194960 CTGTTATCGTAGAACGCTTGAACACGAGCAGTTGGGCTTGGTTTCAACGAGCTAATG 2194901
QY 1448 GAGAATATGAAGACCTTATTAAGAGCTGTTGATTCATTCCTCAAAAAGTTACACGCTATG 1507
DB 2194900 GCGAATGGTAAACATGATTGACGCTGGTATTGTTGATCCAAACAAAGTTACTCGTCCG 2194841
QY 1508 CATTACAAAATGACAGCATCAGTAGCTCTTACTTCTAACTACAGAAATGCGCTATTGCTG 1567
DB 2194840 CATTACAAAACGCTTCACTGTTGCTCTTCTTAACTACCGAAGCTGTTGTAGCAG 2194781
QY 1568 AAAAACCCAGAACCTTAAAAAGATATGCTATGCTGGCGGTGATGGTGGTATGGGTTG 1627
DB 2194780 ATCAACCCAGATGAAAAACCGCCAGCAGCGCCAGATATGGAATGGGCGCATGGCG 2194721

QY 1628 GTATG 1632
 Db 2194720 GTATG 2194716
 RESULT 9
 ID ABA03041/C
 AB 03041 standard; DNA; 2944528 BP.
 XX
 AC ABA03041;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes EGD-e genome sequence.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease; ds.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR011118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Prangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Bolland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, de Pablos B, Wehland J, Kaerst U, Encian K, Hauf J;
 PI Rose M, Voss H;
 XX
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 PS Claim 1; SEQ ID No 1; 192pp; French.
 XX
 CC The present sequence is the genome sequence of Listeria monocytogenes
 CC EGD-e. This sequence and fragments of this sequence are useful for
 CC selecting probes and primers for detecting genes in L. monocytogenes and
 CC related organisms, and to study genetic polymorphisms and other genomes.
 CC Proteins (AB47297-AB47314) expressed from the present sequence are
 CC useful for raising specific antibodies, identification of L.
 CC monocytogenes and related organisms and for biosynthesis and
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
 CC proteins encoded by it are also useful for selecting compounds that
 CC regulate gene expression and cell replication and modulate L.
 CC monocytogenes-related diseases. In addition, this sequence and proteins
 CC encoded by it are useful in pharmaceutical and vaccine compositions for
 CC the treatment or prevention of infections by L. monocytogenes and related
 CC organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
 XX
 Query Match 43.9%; Score 722.6; DB 24; Length 2944528;
 Best Local Similarity 66.2%; Pred. No. 9.9e-156;
 Matches 1075; Conservative 0; Mismatches 544; Indels 6; Gaps 2;
 8 CTAAGAAATCCTTTTGTATGCTAAGCCGCGGAAACCTTCACGAGGTAGATAAAC 67

Db 2149308 CAAAAGATATTAATTTAGTGAAGATGCTCGTGTGCTGCTAGTGTGACCAAC 2149249
 QY 68 TTGCAAAATGCTGTTAAAGTAACACTTGGACCTTAAGCCGCTAATGCTGTTATTGAAAGT 127
 Db 2149248 TAGCAAAACGCAAGTAAAGTAAACGCTTTGGCCCAAAAGGTCGTAATGTTGTTTAGAGAAGA 2149189
 QY 128 CTTTGTGTTCCCGAGTTATTACAAAAGATGCTGCTATCTGTGCAAAAAGAAATGAACTTG 187
 Db 2149188 AATTGCGTTCTCGTTAAATTACAAATGATGGGTACAAATTCGCAAGAAATGAAATTAG 2149129
 QY 188 AAGATAAGTTTGAATAATATGGCGCTCAAATGTTTAAAGAGTAGTCCCAAAACTAGCG 247
 Db 2149128 AAGACCCATTTTGAAATATGCGGAGCAAACTCTGTATCTGAAGTTGCTTCTTAAACCAATG 2149069
 QY 248 ATATTGCTGTGATGAACTACCAACAGCAACAGTCTCTTGCACAGCTATTATTTCGTGAAG 307
 Db 2149068 ATGTTGCTGGGACGGAACCTACCAACAGCTACCGTTTTAGCGCAAGCAATGATTCAAGAAG 2149009
 QY 308 GTGTAAACCTGTAGCAGCTGCTGTAATCTATGGCCATTAAACGTGGCATAGATAAAG 367
 Db 2149008 GCTTAAAAAATGTAAACAGCTGGAGCAACCCAGTAGGCGTTTCGCCGCGGTATCGAAAAG 2148949
 QY 368 CTGTTGTTGCTGTTTACTTAAAGAACTAAGCCACATTAACAAACCTTCTCGTGACCAAAAG 427
 Db 2148948 CCGTAGCAACAGCTATCGAAGAAATTAAGCTATTTCTAAACCAATTTGAAAGCAAGAGT 2148889
 QY 428 AATAGCTCAAGTTGGAAACCATTTCTGCAAACTCTGATACAACTAGTAAATATCATAG 487
 Db 2148888 CTATCGCTCAAGTTGCTGCTATTCTTC---TGGTGTAGTGAAGAGTTGGTAAATTAATCG 2148832
 QY 488 CTGAAGCTATGCTTAAAGTTGAAAAGGAGGTGTTATCACAGTTTGAGGAAGCTAAAGTC 547
 Db 2148831 CAGAAGCAATGGAACCGGTTGGTAACGACGCGTTTACCTATTGAGAATTCGAAAGTT 2148772
 QY 548 TTGAAAACCTACATAGATGTTGTTGAAGAAATGAAGTTTGACCGGTGCTCTCTCCAT 607
 Db 2148771 TTGCAACAGAAATAGACGTAGTAGAAGTATGCAATTCGACCGCGGTACACTAGCCCTT 2148712
 QY 608 ACTTTGTAACCTAATCTCTGAGAAATGTTTGTGTAACCTTGATAACCTTATATCCTTTGTA 667
 Db 2148711 ACATGTTAACTGATTCGATAAAATGGAAGCTGTTCTTTGAAAAACCATACATTTTAATTA 2148652
 QY 668 ATGAGAAAAGATTTACTAGCATGAAGACATGCTACCAATCTTAGACAACAGTGTCTAAG 727
 Db 2148651 CAGCAAAAAAATCAACAATTTCAAGAAATCTTACAGTTTTAGAACAGTGTCTCAAC 2148592
 QY 728 TAAACCGTCCACTCTTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCACACCTTG 787
 Db 2148591 AAGTCTGTCCTCAATGTTAATCATTGCGGAAGATGTTGAAGGGGAAGCTCAAGCAACTCTTG 2148532
 QY 788 TAGTCAATAGCTCCGTGGAGCACTCCAGTTGTAGCGTGAAGAAAGCTCTGTTTGGTG 847
 Db 2148531 TACTAAACAAACTTCGCGGAACATTTAAACGTAGTAGCGGTGAAGAGCTCTGTTTGGTG 2148472
 QY 848 AACCCGTGAAGCTATGCTGAAGATATGCTATCTCTTACTGGAGGAGAAGCAATATTG 907
 Db 2148471 ATCGTGAAGCAATGCTAGAGATATGCTATTTTAAACAGCGGCAAGTCAATACAG 2148412
 QY 908 AAGATCGTGTATTAAGCTTTGAAAATGTAAGCTTTGCTCTTCTTTAGGAACAGCTAAACGTG 967
 Db 2148411 AAGACCTAGGATTAGAACTAAAAACAGCTACTGTTGATCACTTGGAAACAGCTAACAAAG 2148352
 QY 968 TAGTTATTGACAAAGAAATTAATCTATCTGTTGATGCTGCGGAAATCAGAAATATTA 1027
 Db 2148351 TAGTCGTAACAAAAAGATGATACAACTCGTAGAAGGAGGAGCGGATTTCCACACAAATTA 2148292
 QY 1028 AAGCTCGAGTTAAACAAATTCGTCACAAATTTGAAGAAACAGCTCAGATTATGATCGT 1087
 Db 2148291 GCGTCTCGGTAAACCAAAATCGTGGCAAAATGGAAGAAATCTTCTGAAATTTGATAGAG 2148232
 QY 1088 AAAAATCTCAAGAAAGCTCTTGCAAAACCTTTGTTGGAGTAGTGTGTTATCCATGTTGGAG 1147

Db 2148231 AAAAATACAGAACCGTTAGCAAAAATTCAGGTGGAGTAGCTGTTTAAAGTCGGCG 2148172
QY 1148 CTGCTACTGAACCTGAAATCAAGAGAGAGAGATCGTGTAGAGATGCTCTAAATGCAG 1207
Db 2148171 CTGCACTGAACAGAGCTAAAGAGAGCTAAATAGTATTGAGATGCACTTAACTCTTA 2148112
QY 1208 CAAAGCTCGGTGTAAGAGAGTATTGTCCTCGTGGTGGTACTGCTTTTGTCCGTCCA 1267
Db 2148111 CTGCGCAGCTGTAGAGAGAGGTATCGTAGCTGGTGGTACTGCTCCCTTTGTAAGTATT 2148052
QY 1268 TTAAGTCTCTGATGATATTAACCTGCTGATGATGATGATGATGATGATGATGATGAT 1327
Db 2148051 ACAATAAGTAGCAGCACTTGA--GCAAGAGGCGACGTAGAGAGAGGTATCAACATTG 2147995
QY 1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTCGTCAAATTCGTCAAATTCGTCAAATTCGTC 1387
Db 2147994 TCGTCGTTCTCTAGAGAGCAGTTGCGCAATTCGCACATACGCTGAGCTTGAAGGTT 2147935
QY 1389 CTATTGTTGTAGAAAAGTTCGTGAACCAAGAGATGTTTGGATTTAATGCTGCATCAG 1447
Db 2147934 CGGTTATCGTTGAACGCTTGAACACAGAGCAGTTGGTGTGGTTTCAATGCAGCAACG 2147875
QY 1448 GAGAAATGAGACCTTATAAGTGTGTGATGATGATGATGATGATGATGATGATGATGATG 1507
Db 2147874 GCGAATGGGTAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2147815
QY 1508 CATTACAAAATGCAGATCAGTAGCCTCCTTACTTCTACTACAGAAATGCTGCTGATGCTG 1567
Db 2147814 CTCTACAAATGCATCATCTGTTGAGCGCTTCTATTACTACAGAGCAGCTGATGAG 2147755
QY 1568 AAAACCCAGAACCTTAAAGATATGCTATGCTGCGCGGTGGTATGGTGGTGGTGGTGGTGG 1627
Db 2147754 ACAACCCAGAACCGCCAGCAGCTGTTCTCATATGGAATGGTGGTGGTGGTGGTGGTGG 2147695
QY 1628 STATG 1632
Db 2147694 GCATG 2147690

RESULT 10

AAT42063

ID AAT42063 standard; DNA; 1830121 BP.

AC AAT42063;

AC AAT42063;

DT 14-SEP-1999 (first entry)

XX 14-SEP-1999 (first entry)

XX Haemophilus influenzae complete genome sequence.

DE Genome; bacterium; Haemophilus influenzae; computer readable medium;

KW expression modulating fragment; regulation; gene expression; vector;

KW organism; open reading frame; ORF; ds.

XX Haemophilus influenzae.

OS

PN WO9633276-A1.

XX 24-OCT-1996.

XX 22-APR-1996; 96WO-US05320.

XX 07-JUN-1995; 95US-0487429.

XX 21-APR-1995; 95US-0426787.

XX 07-JUN-1995; 95US-0476102.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UJO) UNIV JOHNS HOPKINS.

XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

XX WPI; 1996-485782/48.

XX Haemophilus influenzae Rd genome recorded on computer readable

PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching

XX Claim 1; Page 77.2-77.1091; 1291pp; English.

CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H-influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.

XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 43.8%; Score 721; DB 17; Length 1830121;
Best Local Similarity 65.7%; Pred. No. 2e-155;
Matches 1083; Conservative 0; Mismatches 560; Indels 6; Gaps 2;

QY 1 ATGGCTTCTTAAAGAAATCCTTTTGTATGCTAAAGCCCGTCAAAACCTTTTCAAGAGGTGA 60
Db 565350 ATGGCAGCAAAAGACGTAAATTTGTAATGACGACGCGTAAATGCTTAAAGCGGTG 565409

QY 61 GATAAATCTTGAATGCTTTAAAGTAACACTTGGACCTTAAAGCCGCTAATGCTGTTATT 120
Db 565410 AATGATTATGCGGATGCACTAAAGTAACCTTTGGCCCGAAAGGCTGTCATGTAATTTA 565469

QY 121 GAAAGTCTTTTGGTTCCTCCAGTATTACAAAAGATGGTGTATCTGTGCAAAAGAAAT 180
Db 565470 GATAAATCAATTTGGCGCACTATCACTAAGACGGTGTCTGTCTGCTGGAATC 565529

QY 181 GAACTTGAAGATAAGTTTGAAATATGGCGCTCAATGTTTAAAGAGTAGTCTCCAAA 240
Db 565530 GAATTAGAGATAAATTCGAAACATGCGGCGCACAAATGCTGAAAGAGTGGCATCTAA 565589

QY 241 ACTAGGATATTGCTGGTGTATGGAATCAACAGCAACAGTCTTGCACAAGCTATTAT 300
Db 565590 GCAAAATGCTGTCAGGTGACGTACAACTACTGCAACTGTACTTGGCGAAGCTATCGTA 565649

QY 301 CGTGAAGGTGTAATACTTGTAGCAGTGTCTGATGCTTAAATGCTTAAAGAGTGGCATA 360
Db 565650 AATGAAGCTTGAAGCAGTAGTCTGAGGTATGAATCCAAATGATTTAAACGTGTTAT 565709

QY 361 GATAAAGCTGTTGTTGCTGTTTAAAGAACTAAGCGACATTTACAAAGCTTACTCTGAC 420
Db 565710 GATAAAGCAGTAAGTGGGTGCTTCTGAACTTAAATAATTTATCTAAACCTTGTGAACA 565769

QY 421 CAAAAGAAATAGCTCAAGTTGGAACCACTTCTGAAACTCTGATACACAATAGGTAAAT 480
Db 565770 GCAAAAGAAATGAACAAGTAGGACTATTTCTGCAAACTCTGACAGCATTTGGGTCAA 565829

QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGAAAAGGAGGTGTTATCACAGTTCGAGAGCT 540
Db 565830 TTAATCTCACAAGCAATGGAAGAGTGGCAGAGAGGTGTTACTCTTGAAGATGTT 565889

QY 541 AAAGCTCTTGAAACTTACATTAGATGTTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db 565890 ACGGCTCTTGAAGATGAATTAGATGTTTGAAGCATGCAATAGATAATGTTTACCTT 565949

QY 601 TCTCCATCTTGTAACTTAATCTCAGAAAATGGTTTGAAGTGTGTAACCTTGAACCTTATATC 660
Db 565950 TCTCCATTTTCAACAACAAACCAAACTGCAACCGTTGAATTAGATAATCCATATCTT 566009

QY 661 CTTTCTTAATGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTT 720
Db 566010 CTTCTTGTAGATAAAAAAATCTTAACATCTGTAATTAATCTCCGCTGTAGAGCGTT 566069

QY 721 GCTAAGTAAACCGTCCACTCTTATTTGCTGAAGAGCTGAGAAGGTGAAGCACTTGA 780

QY 1 ATGGCTTCAAGAAATCCCTTTTGTGCTTAAGCCCGTGAAGAACTTTTACAGAGGTGA 60
Db |||||
673 ATGGCAGCAAAAATATTAATATAAGAGAGCGCCAGAAAAAATCCATAAGAGTT 732
QY 61 GATAAATTTGCAATCTGTTAAAGTAAACACTTGGACCTTAAGCCGCTTAATGCTTAT 120
Db |||||
733 AAAACCTTTGCAAGACTGTAAAGTAACTTAGTCCCTAAAGCCGCTCATGTGGTTATC 792
QY 121 GAAAGCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTGTCARAAAGAAAT 180
Db |||||
793 GATAAAGCTTTGGTTCTCTCAAGTTACCAAGATGGCGTAACTGTGCGTAAAGAAAT 852
QY 181 GAATTTGAAGATAAGTTTGAATATGGCGCTCAAAATGGTTAAAGAGTAGTCCCAAA 240
Db |||||
853 GAGCTCGAAGACAAGCATGAGAACATGGGAGCTCAAAATGGTTAAAGAGTCTGCTAGCAA 912
QY 241 ACTAGCGATTTGCTGGTGAATGAACCTACACAGCAACAGTCTTTCGACAGCTATTTAT 300
Db |||||
913 ACTGCGAGTAAAGCTGGTGAATGAACCTACACAGCTACTGTTCTTCGAGAAGCTATCTAC 972
QY 301 CETGAAGGTGTAAAACTTGTAGCAGCTGGTGGTAAATCCCTATGGCCATTAAACGTGGCATA 360
Db |||||
973 AGTGAAGATTGAGAACTGAATCGAGCGGCCAATCTTAGACCTCAAGAGAGCAAT 1032
QY 361 GATAAGCTTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCTGAC 420
Db |||||
1033 GATAAGCGATAAAGCTGGTGTGATGAATCAAAAAAATAGTAAACCCGCTACAAAT 1092
QY 421 CAAAAGAAATAGCTCAAGTTGGAACCAATTCGCAAACTCTGATACAACTAGTAAAT 480
Db |||||
1093 CACAAGAAATAGCTCAAGTAGCGACTATTTCTGCAAAATAGTATGCTGAAATCGGTAAT 1152
QY 481 ATCATAGCTGAAGCTATGCTAAAGTTGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db |||||
1153 CTATTCGCCGAGCCATGAAAAGTTGCAAAAACGCTCTATTAAGTTGAGAGCT 1212
QY 541 AAAGCTTTGAACTACATTAGATGTTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
Db |||||
1213 AAAGTTTTCGAACTCTCCGACGTTGTGCAAGGTATGAATTTCAACCGCGATACCTA 1272
QY 601 TCTCCATCTTTGTAATCTCTGAAAAATGGTTTGAACCTGATTAACCTTATATC 660
Db |||||
1273 TCAGCTACTTCTTCAAAATCTGAAACCAAGAAATGTTTGAAGAAAGCTCTCGTG 1332
QY 661 CTTTGTATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
Db |||||
1333 CTATATCATGATAAAAAATTTCCGGAATCAAGATTTCTACCACTTTTACCAAGTA 1392
QY 721 GCTAAGTAAACCGTCCACTCCTTATATTGTGTAAGACGTAGAAGGTGAAGCACTTGA 780
Db |||||
1393 GCAGAATCAGGACGTCCTTACTTATCAATGCTGAAGATATCGAAGGAGAGCTTTAGCT 1452
QY 781 ACACCTGTAGTCAATAAGCTCGTGGAGCACTCCAGTTGTAGCCGTAAAGCTCTGCT 840
Db |||||
1453 ACTTTAGTAGTAAACAGACTACGTCGTGAATTCAGAGTGTGTCAGTAAAGCTCTGGA 1512
QY 841 TTTGTGAACCGCGTAAAGCTATGCTTGAAGATATTGCTTATCTTACTCGAGGAGAGCA 900
Db |||||
1513 TTTGTGATACAAACAAAGCTATGTTAGAGACATCGCTATTTTAACTGGTGGTCACTC 1572
QY 901 ATATTTGAAGATCGTGGTATAAGCTTGAAGTGAAGCTTCTTCTTTAGGAACAGCT 960
Db |||||
1573 ATCAGCGAAGAGCTTGGCATGAAGCTTGAGAACACAACTCTAGCTATGTTAGGAAAGCT 1632
QY 961 AAACGTGTAGTTATTGACAAAGAAATCTACTATCTGTTGATGCTGGTGAAGAAATCAGAA 1020
Db |||||
1633 AAAAAGTCACTGTTTCCAAAGAGATACAAATTTGTTGAAGGTCTTTGGCAGCAAGAA 1692
QY 1021 GATATTAAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTAT 1080
Db |||||
1693 GATATTTGAATCTCGTGGCAAGATATCAAAAAACAAATCGAAGACAGTACTTCTGATTAC 1752
QY 1081 GATCGTGAAGAACTTCAAGAACGCTTTCGAAAACTTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db 1753 GACAAAGAAAACTCCAAGAACGTTTAGCTTAACCTTCCGAGCGGTAGCTGTAATCCGT 1812
QY 1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
Db 1813 GTAGGAGCTGCTACAGAAATCGAAATGAAAGAGAAAAAGACAGAGTAGATGCTCAG 1872
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCTCGTGGTGTGCTGCTTTGTC 1260
Db 1873 CATGCAACTTCTGCTGAGTTGAAGAGGTATTCTACCTGGCGGTGCTAGCTTTAGTT 1932
QY 1261 CGCTCCATTAAAGTCCCTGATGATAT-----TAAACCTGCTGATGATGAACCTTGT 1314
Db 1933 CGCTGCATCCCTACTTTAGAGCTTTTCAATTCCTATTCTTACAAATGAAGATGAGCAATC 1992
QY 1315 GGAATTAATATCATCGTCTGCTTCTTGAAGAGCTTTAGTCAAAATTTGCTGCAAAATGCT 1374
Db 1993 GGAGCAGCTATTGTTCTCAAGCATTTATCGCTCCATTAAAGCAAAATTCAGCCAAATGCT 2052
QY 1375 GGCTATGAAGGTCTATTGTTGTAGAAAAAAGTTTCGTGAACCAAAAGATGTTTGGATT 1434
Db 2053 GGTAAGAGAGCGCTATCATCTGTCAAGAGTCTTTCTGCTCTCTAGCGAAGCTAT 2112
QY 1435 AATGCTGCATCAGGAGATATGAAGACCTTATTAAAGCTGGTGTCTATTGATCCTAAAAA 1494
Db 2113 GATGCTTTACGCGATGCTTTACACCGACATGATTTGAGCGAGGAATTTCTCGATCCAACTAA 2172
QY 1495 GTTACAGCTATTGCTATTACAAATGCGAGCATCAGTAGCCTCTTACTTCTTACTACAGAA 1554
Db 2173 GTTACAGCTGTGCTTTAGAAAAAGCGAGCTTCTGTAGCTGGGCTTCTATTAAACAGAA 2232
QY 1555 TCGCTATTGCTGAAAAACAGAACCTAAA 1584
Db 2233 GCTTTAATTCGGATATTTCTCGAAGAGAA 2262

RESULT 12

AAQ13137

ID AAQ13137 standard; DNA; 2223 BP.

XX AAQ13137;

XX 25-MAR-2003 (updated)

DT 17-DEC-2001 (updated)

DT 22-OCT-1991 (first entry)

XX Hyp operon.

XX Antibodies; heat shock; hypersensitive; allergen; HSP60; GroEL; ss.

XX Chlamydia trachomatis serovar A.

Key Location/Qualifiers

FH CDS 35..343

FT /*tag= a

FT /label= hypA

FT 381..2015

FT /*tag= b

FT /label= hypB

FT 23..26

FT /*tag= c

FT 35..38

FT /*tag= d

FT 2046..2072

FT /*tag= e

FT terminator

FT /note=" dyad symmetry; rho-independent"

XX USN7531317-N.

XX 09-JUL-1991.

XX 31-MAY-1990;

XX 90US-0143560.

XX

31-MAY-1990; 90US-0143560.
(USSH) NAT INST OF HEALTH.
WPI: 1991-245693/33.
P-PSDB; AARI13336, AARI13337.
DNA encoding HypA and HypB Chlamydia proteins - used to develop
prods. for detection of and vaccines against Chlamydia infection.
Disclosure; Fig 7; 51pp; English.
The sequence was obt'd. from clone pTA571 isolated from a DNA library
prep'd. from C. trachomatis genomic DNA. The two ORFs encode HypA
and HypB hypersensitivity proteins. The DNA can be used to produce
recombinant proteins or to design probes for the detection of
Chlamydial infection. The Hyp proteins can be used to raise a cell-
antibodies and in vaccines. The Hyp B protein elicits a cell-
mediated immune response so can be used as a skin test antigen.
See also AAQ13136.
(Note: Revised entry submitted to correct the patent number format of
US Government-owned NRIIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/nris.us.html.)
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 2223 BP; 745 A; 408 C; 489 G; 581 T; 0 other;
Query Match 43.1%; Score 710.4; DB 12; Length 2223;
Best Local Similarity 65.9%; Pred. No. 6.9e-154;
Matches 1048; Conservative 0; Mismatches 536; Indels 6; Gaps 1;
1 ATGGCTTCTAAAGAAATCCCTTTTGTGATGCTAAAGCCCGCTGAAATCTTCACGAGGTGTA 60
381 ATGGCTCGTAAAAACATTAATAACAAGAGAGCCAGAAAGAAAATTCAAAAGAGGT 440
61 GATAAATCTGCAATCTGTTAAAGTAACTTGGACCTAAAGCCCGCTTAATGCTGTTATT 120
441 AAGACTTTAGCTGAAGCTGTAAGATGCTTACCTAGGGCTTAAGGACGACATGTTGTCATA 500
121 GAAAGCTTTTGGTCCCGAGTTATTACAAAGATGGTGTATCTGTTGCAAAAGAAAT 180
501 GATAAAGCTTCGATCCCTCAAGTAACTAAAGATGGTGTATCCGTTGCAAGAGATT 560
181 GAACCTGAAGATAGTTTGAATAATATGGGCTCAATATGTTAAAGATGCTCCAAA 240
561 GAGCTTCCGACAAACATGAAATATGGGCTCAATATGTTCAAGAGATGCTCCAGCAA 620
241 ACTAGCGATATGCTGATGGAATACACAGCAACAGCTCTTGCACAGCTATTAT 300
621 ACTGCTGACAAAGCTGGAGCGGAATACACAGCTACTGTTCTGCTGAAGCTATCTAT 680
301 CGTCAAGGTGTAAGCTTGTAGCAGCTGGTCTATCTATGCGCATTTAAACGTTGGCATA 360
681 ACAGAGATTAACGATGTAACAGCTGAGCAATCAATGGACCTCAACAGAGATT 740
361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGCATTAACAGCTTACTGTCGAC 420
741 GATAAAGCTGTTAAGTTGTTGTTGATCAATCAAAATAATCAGCAAGCTGTTTCAGCAT 800
421 CAAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAACTAGGTAAAT 480
801 CATAAAGAAATAGCTCAAGTTGCAATTTCTGCTAATAATGATGCAAGAAATCGGGAAT 860
481 ATCATAGCTGAAGCTATCGCTAAAGTTCGAAAGGAGGTGTTATCAGAGTTGAGGAAGCT 540
861 CTGATGCTGAAGCAATCGAGAAAGTTGTTAAACGCTCTATCAGCTGTTGAAGAGCA 920
541 AAAGCTCTTGAACATCAATTAGATGTTGTTGAAGGAATCAAGTTTACCGTGGCTACCTC 600
921 AAAGATTTGAACCGTTTGTGATGTTGTTGAAGGAATGAATTTCAATAGAGGTGTACCTC 980
601 TCTCCATATCTTGTAACTAATCTCAGAAATGGTTTGTGAACCTTGATACCCCTTATATC 660

981 TCTAGCTACTTCGCAACAAATCCAGAAATCTCAAGAAATGTGATTAGAACGCTTTGGTT 1040
661 CTTTGTAAATGAGAAAGAAATGTTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTT 720
1041 CTAATCTACGATAAGAAATTTCTGGGATCAAGATTTCCCTCTATTTTACAAAGATT 1100
721 GCTAAAGTAAACCGTCCACTCCCTTATTATTGCTGAAGAGCTAGAGGTGAAGCATTGCA 780
1101 GCTGAATCCGGCTGCTCTCTTATTATAGCAGAGACATTTGAAGCGAAGCTTTAGCT 1160
781 ACATTTGTAGTCAATAAGCTCCGTGAGCAGCTCCAAAGTTGTAGCCGTAAAGCTCTGGT 840
1161 ACTTTGGTCTGTAACAGAAATTCGTGAGGATTCGGGTTTTCGCGAGTTAAAGCTCCAGGC 1220
841 TTTGGTGAACCGCTGAAGCTATGCTTGAAGATTTGCTATCTTACTCTGAGAGAGCA 900
1221 TTTGGAGATGAAGAAAGCTATGTTGGAAGACATCGCTATCTTAACTGCGGTCAACTC 1280
901 ATATTTGAAGATCGTGTATATAAGCTTGAAGATGTAAGCTTGTCTCTTTAGGAAACAGCT 960
1281 ATTAGCGAGAGTTGGCATGAATTTAGAAACGCTAACTTAGCTATGTTAGTAAAGCT 1340
961 AAACGTGTAGTTATGACAAAGAAATPACTATCTGTTGATGCTGCTGCTGCTGCTGCTG 1020
1341 AAAAAAGTTATCGTTTCTTAAAGAAAGACACGACCATCGTGAAGGAATGCGTGAAGAAAG 1400
1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTTGAAGAAACAGCTCAGATTAT 1080
1401 GCTTTGAAGCTCGTTTCGAAAGCATCAAAAAACAAATTTGAAGACAGCTCTTCTGATAC 1460
1081 GATCGTGAAGAACTTCAAGAAAGCTTCTTGAAGAACTTCTGCTGAGTAGCTGTTATCCAT 1140
1461 GATAAGAAAGAACTTCAAGAGCTTCTGCTGTAAGCTCTCTGCTGAGTAGCAGTCACTCGC 1520
1141 GTTGGAGCTCTACTGAACTGAAATGAAAGAAAGAGATCTGTAGAGATGCTCTA 1200
1521 GTTGGAGCTCAACAGAGATTGAGATGAAAGAAAGAAAGATCGTGTAGACGATGCTCAA 1580
1201 AATGCAACAGAGCTGCGGTTGAAGAGGTTATGCTCCCTGCTGCTGCTGCTGCTGCTGCT 1260
1581 CATGCTACAATCGCTGCTGTTGAAGAGGAATTTCTTCTGCTGAGGAGAAACAGCAATTAATC 1640
1261 CGCTCCATTAAGCTCCCTGATGATTAATAAAGCTG-----CTGATGATGATGAACCTTGC 1314
1641 CGTTGTAATCCCTACTCTTGAAGCTTCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1700
1315 GGACTTAATATCACTCCGCTCTCTTCTTGAAGAGCTTTTACGTCAAATTTGCTGCAAAATGCT 1374
1701 GGAGCTCGCATTTGTTGAAAGCTCTTTTCCGCTCTCTTGAAGCAAAATTTGCTGCAAAAGCA 1760
1375 GGCTATGAAGCTTCTATTTGTTAGAAAGAGTTGCTGTAACAAAGAGTGGTTTGGATT 1434
1761 GGAAGAGAGGTGCTATCATCTTCCAAAGTTATGCTCCGCTTCTGCGAAGCAAGGATAT 1820
1435 AATGCTGATCAGGAGAAATGAAAGCTTATTAAGCTGGGTGCTCATGATCTCTTAAAGAA 1494
1821 GATGATTTGCTGATGATACAGATATGCTGAGCTGCTGATTTTAGATCTCTGCTTAA 1880
1495 GTTACAGCTATTGCTATTAACAAATGCAAGCATCAGTAGCTCTCTTACTTCTTAACTACAG 1554
1881 GTTACCGCTTCTGCTTTTGAAGAGCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1940
1555 TGCGCTATTGCTGAAGAAACAGAAACCTTAA 1584
1941 GCTCTCATTTGAGAGATTCCAGAAAGAAAA 1970

RESULT 13
ABA92787
ID ABA92787 standard; DNA; 640681 BP.
XX
AC ABA92787;

XX 27-MAR-2002 (first entry)
XX Buchnera sp. genomic DNA SEQ ID NO:1.
XX Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
XX circular; ds.
XX Buchnera sp.
XX JP2001292771-A.
XX 23-OCT-2001.
XX 07-APR-2000; 2000JP-0107160.
XX 07-APR-2000; 2000JP-0107160.
XX (RIKA) RIKAGAKU KENKYUSHO.
XX WPI; 2002-126043/17.
XX A genomic DNA of cockroach-symbiotic bacterium -
XX Claim 1; Page 16-230; 237pp; Japanese.
XX The present invention describes a gene (I) derived from Buchnera sp.
XX containing the DNA (a) or (b), (a) has a fully defined base pair
XX sequence selected from a table of sequences found in the Buchnera sp.
XX genomic DNA of ABA92787 given in the specification or is a DNA selected
XX from complementary DNA sequences, and (b) is a DNA which hybridises with
XX the DNA (a) and encodes a protein. Also described are: (1) a recombinant
XX vector (II) containing (I); (2) a transformant (III) containing (II);
XX (3) a genomic DNA of Buchnera sp. containing the sequence given in
XX ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
XX (d), (c) is a DNA containing a fully defined sequence given in
XX or ABB92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
XX method for the preparation of a protein in which (III) is cultured and
XX the expression protein of the objective protein is collected from the
XX resultant culture. The DNA is useful for developing agricultural
XX chemicals for exterminating cockroaches. The present sequence represents
XX the specifically claimed Buchnera sp. genomic DNA sequence, from the
XX present invention.
XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
Query Match 43.1%; Score 710; DB 24; Length 640681;
Best Local Similarity 64.5%; Pred. No. 4.9e-153;
Matches 1061; Conservative 0; Mismatches 585; Indels 0; Gaps 0;
QY 1 ATGGCTTCTAAGAAATCCCTTTTGGATGCTAAAGCCCGTGAAAAAATCTTTCAGGAGTGTA 60
DB 18715 ATGGCCGCTAAGAGATGTAATTTTGGAAATGAAGCCCGCATTAATGCTTGGAGTT 18774
QY 61 GATAAAGCTGCAATGCTGTTAAAGTACACTTGGACCTAAAGCCGCTAAAGTTCGTTAT 120
DB 18775 AATGATTAGCAGATGCGAGTAAAGTACTTTAGCACCACCAAGGAGTAAATGCTGTA 18834
QY 121 GAAAGTCTTTGGTTCCTCCAGTTTACAAAGATGTTGATCTGTTGCAAAAGAAAT 180
DB 18835 GATAAATCTTTGGAGCCTAGTATTAATAAGATGTTGATCTCGTAGCCCGTGAAT 18894
QY 181 GAACCTTGAAGATAAGTTTGAATAATGGCGCTCAATGGTTTAAAGAAAGTAGCTCCCAA 240
DB 18895 GAAATTAGAAGATAAATTCGAAACATGGAGCTCAATGGTAAAGAAAGTTGCATCAAA 18954
QY 241 ACTAGCGATATGCTGGTGAAGGAACTACACAGCAAGCTCTTGGCAAGCTATTAT 300
DB 18955 GCAACAGATGAGCAGGATGGTGGTACCAACACAGCAACATTAATAGCAATATAGTA 19014
QY 301 CGTGAAGGTGAAATCTTGTAGCAGCTGCTGATCTGATCTGATCTGATCTGATCTGAT 360
DB 19015 AATGAGGTTTAAAGAGTAGTAGCAGCTGATGATGATGATGATGATGATGATGAT 19074

QY 361 GATAAAGCTGTTGCTGTTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
DB 19075 GATAAAGCTGTTTACTAGTCTGTAGAGAAATTAACAATTTATCTGTACATGTTCTGAT 19134
QY 421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAAACAATAGGTAAT 480
DB 19135 TCTAAAGCAATTACACAAGTTGGTACTATTCTGCAANTGCAGATGAAAGTTGGTTCT 19194
QY 481 ATCATAGCTGAAGCTATGCTAAAGTTGGAAGAGGTTGTTATCAAGTTGAGGAGGCT 540
DB 19195 TTAATTCGAGAAGCAATGGAAAAAGTTGGTAATGACGAGTTATTACAGTAGAAGAGGT 19254
QY 541 AAAGGCTTTGAAACTACATTAGATGTTGTTGAAGGAATGAAGTTTCACCGTCTACCTC 600
DB 19255 ACAGGTTTACAGATGAACCTGAAGTTGCAAGGATGCAATTTGATCGTGGTTATCTA 19314
QY 601 TCTCCATCTTTGTAACCTAATCCTGAGAAAAATGGTTTGTCAACTTGATAACCCCTTATATC 660
DB 19315 TCTCCATATTTTATCAATAAACCCAGAAACAGGTATTGTTGAAATTAGAAACCCATATAT 19374
QY 661 CTTTGTATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAAGTT 720
DB 19375 TTAATGGCTGATAAAAAAATATCTAATGTTCTGTAATGTTACCAATATTAGAAATCTGTT 19434
QY 721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAGACGTAGAGGTGAGGACTTGCA 780
DB 19435 GCAAAATCAGGAAAAACCATTAATTAATTTCTGAAAGTTTAGAGGGTGAAGCTTTAGCA 19494
QY 781 ACACCTTTAGTCAATTAAGCTCCGTGAGCACTCCAAGTTGTAGCCCTAAAAGCTCTGCT 840
DB 19495 ACTTTAGTAGTTAATTAATCAATGAGGAAATGTAAAAGTCGACAGCTAAAGCACTGGA 19554
QY 841 TTTGTTGAGCCGCTAAAGCTATGCTGAAAGATTTGCTTCTTTAGGACAGCT 960
DB 19555 TTTGTTGATCGCTGTAAGCAATGTTACAGAGATTTTCAATTTCTTACTGCTGCTGTT 19614
QY 901 ATATTGTAAGATCGTGGTAAAGCTTGAAATGTAAGCTTCTTCTTTAGGACAGCT 960
DB 19615 ATCTCTGAAGATTTAGCTATGGAATTAAGAAATCTCTTTAGAGATTTAGGACAGCA 19674
QY 961 AAACGCTAGTTATTGACAAAGAAATATCTATCTGTTGATGTTGTTGTTGTTGTTGTTGTT 1020
DB 19675 AAACGCTGTTTATTAGCAAGACACTACAATTTATTGTTGTTGTTGTTGTTGTTGTTGTT 19734
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTAT 1080
DB 19735 TCCATTCAAGTCTGTTATTAGTCAATCCGACAAAGAAATTTCAAGAAAGCTACTTCTGATTAT 19794
QY 1081 GATCGTGAAGAACTTCAAGAACGCTTTCGAAACCTTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
DB 19795 GATAAAGAAATTAATTAATGAAACGCTTAGCTAACTATCAGGCGGTTGTCAGTACTAAA 19854
QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAGAGAGAGAGGATCGTTGTTAGAGATGCTCTA 1200
DB 19855 GTAGTGGCTACAGAAAGTGAATGAAGAAAGAAAGCTCGTTGTTGAAGATCAATTA 19914
QY 1201 AATGCAACAGAGCTGCGGTTGAAGAGGTTATGTCCTCGTTGTTGTTGTTGTTGTTGTTGTT 1260
DB 19915 CATGCTACTCGTCAGCTGTAGAGAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 19974
QY 1261 CGCTCCATTAAAGTCTTGTGATGATTAAACCTGCTGATGATGATGATGATGATGATGATGAT 1320
DB 19975 CGTGTAGCAGGAAAAATAGCTGATTTAGCTGCTCAATTAAGATGATGATGATGATGATGAT 20034
QY 1321 AATATCATCGCTGCTTCTTTGAAGAGCTTTAGCTCAATTTAGCTGCAAAATGCTGCTGCTAT 1380
DB 20035 CGAGTTGCTTTGCGTGAATGAAAGCTCCAATTAGCTCAATTTGTTCTTAATTTCTGTTGAA 20094
QY 1381 GAAAGTCTATTTGTTGTAAGAAAGTTCTGTAACCAAGAGATGTTTGGATTTTAAATGCT 1440
DB 20095 GAACCTTCTGTAGTTACAAACAATGTAAGACCGGAAAGGTAACACTATGTTGTTACAAATGCA 20154

QY 1441 GCATCAGGAGATATGAACACCTTATTAAAGCTGGTGTCTATTCATTCATCCTTAAAAAGTTACA 1500
DB 20155 GCTACTGATGAATATGGTGACATGATAGATTTTGGTATATATAGATCCAACTAAAGTTACA 20214
QY 1501 CQTATTGCAATCAAAATGCAGCATCAGTAGCCTCTTACTTCTAACTACAGAAATCGCT 1560
DB 20215 CGTTCGCTTTACATGATGCTGCTTCTGCTGCTGCTCTAATGATCAACAGAAATGATG 20274
QY 1561 ATTGCTGAAAAACCAACCTTAAAAAGATATGCCCTATGCTGCGGTGATATGGTGGT 1620
DB 20275 GTAACCTGACTGCTTAAAGAGATAAATCTCTGATTTAGTTCTTCTCCAGCAGGAGGA 20334
QY 1621 ATGGGTGGTATGACAGGTATGACTA 1646
DB 20335 ATGGGTGGCATGGTGGGAATGATGA 20360

RESULT 14

ABA90521
ID ABA90521 standard; DNA; 2365589 BP.

XX
AC ABA90521;

XX 16-MAY-2002 (first entry)

XX Genomic sequence of *Lactococcus lactis* IL1403.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.

XX *Lactococcus lactis* IL1403.

XX FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-0004630.

XX 11-APR-2000; 2000FR-0004630.

XX (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification of *Lactococcus*
XX *lactis* and related species -

XX Claim 1; SEQ ID 1; 2504pp; French.

XX The present invention is related to a *Lactococcus lactis* nucleotide
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
XX nucleic acid sequence is useful in the detection and/or amplification of
XX nucleic acid sequence, particularly to identify *Lactococcus lactis* or
XX related species. The proteins of the invention are useful for the
XX biosynthesis or biodegradation of a composition of interest. The
XX invention helps research in lactic bacteria, particularly useful in the
XX production of yogurt and cheese.

XX Note: The sequence data for this patent is based on equivalent patent
XX WO2001/7334 (published 18-OCT-2001) which is available in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

XX Query Match 43.0%; Score 708.4; DB 24; Length 2365589;

XX Best Local Similarity 66.5%; Pred No. 1.7e-152;
XX Matches 1047; Conservative 0; Mismatches 521; Indels 6; Gaps 2;

QY 7 TCTAAAGAAATCCCTTTTGGTATGCTAAAGCCCGTGAATAAATCTTCACGAGGTGATAGATAA 66

DB 398900 TCARAAGATATTAATTTCAAGTGATGCTAGACAGCGATGATCGTGGGATTCATATT 398959

QY 67 CTTCGAATGCTGTTAAAGATAACACTTGGACCTAAAGCCCGTAATGCTGTTATTGAAAG 126

DB 398960 CTTGCTGATACAGTAAAAACAACCCCTTGACAAAAAGSTCGCAATGCTGTTCTTGAAAAA 399019
QY 127 TCTTTTGGTTCCTCCAGTATTATCAAAAGATGGTGTATCTGTTGCAAAAGAAATTTGAACCT 186
DB 399020 TCATATGGATCACCTTTAATTAACAAACGATGGGTAAACGATTCGCAAGAGATGACCTT 399079
QY 187 GAAGATAAGCTTTGAAATATGGCGCTCAAAATGGTTAAAGAACTAGCTCCCAAACTAGC 246
DB 399080 GAAGATCAATTTGAAATATGGAGCTAACTTGTTCAGAACTGCTTCAAAAAACAAT 399139
QY 247 GATATTGCTGGTATGGAACATAACAGCAACAGTCTCTTGCACAAGCTATTTATCTGAA 306
DB 399140 GATATCGCAGGTGACGGTACAAACAGCGACAGTCTTTGACACAAGCTATCGTTCGTA 399199
QY 307 GGTGTAATAACTTTGTAGCAGCTGCTGTAATCCTATGCGCAATTAACGTGTCATAGATAA 366
DB 399200 GGTTTAAAAAATGTTTACCGCAGGTGCAAAATCCAGTTGGTATTGCGCAGGAATGACCT 399259
QY 367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTAACAAAGCCTACTCTGTACCAAAA 426
DB 399260 GCTGCTGAACAGCTGTTGCTCAATTAAGAGATGCAATTCCTGTCCACGATNAATCA 399319
QY 427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGTAATATCATA 486
DB 399320 GCAATTCGCAAGTAGCTACCGTTTCATCACGTAGTGA---AAAGTGGGTGAGTATATT 399376
QY 487 GCTGAAGCTATGCTAAAGTTGGAAGAGGTGCTTATACAGTTGAGGAAGCTTAAAGGT 546
DB 399377 TCTGATGCAATGGAACGCTAGGTTCGACGAGTTATACCAATGAAGAAATCAAAAGG 399436
QY 547 CTTGAAACTACATTAGATGTGTTGAAGGAATGAAGTTGACCGTGTCTCTCTCCA 606
DB 399437 ATGCAAACTGAACTCGATGTTGTTGAAGGAATGCAATTCGACCGTGATATTTGAGTCA 399496
QY 607 TACTTTGTAACTAATCCTGAGAAAATGGTTGTGAACTTGGAATACCTTATATCTTTGT 666
DB 399497 TATATGTTCTTAATACAGAAAAATGGTTGCTGAATTAGATAATCTTATATCTTATT 399556
QY 667 AATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTTGCTTAAA 726
DB 399557 ACCGACAAAAAATCTCAAAACATTTACCGTCTCTTGAAACAAATCTTGAA 399616
QY 727 GTAAACGCTCCACTCTTATTATTGCTGAAGACCTAGAAGTGAAGCACTTGCAACACT 786
DB 399617 ACAATGCTCCACTCTTATTGATGCTGATGATGTTGATGAGAAGCATTCGCAAGCTT 399676
QY 787 GTAGTCAATAGCTCCGTTGGAGCACTCCAGTTGTAGCCGTAAAGCTCTCTGTTTGGT 846
DB 399677 GTTCTTAATAGATTAAGAGGTGCTCTCAATGTTGTAGCTGTAAAGCACCAGGATTTGGT 399736
QY 847 GAACGCGTAAAGCTATGCTTGAAGATATTGCTATCTTACTGAGGAGAGCAATATT 906
DB 399737 GATCGTGAAGCTCAATTTGGAAGATTTGGCTATCTTTGACAGGTGGTACAGTAATTA 399796
QY 907 GAAGATCGTGTATAAAAGCTTGAATAATGTAAGCTTTGCTTCTTTAGGAACAGCTTAAAC 966
DB 399797 GAAGAACTTGTCTGATCTTTAAGATGCTACTCTTGAAGCTTTAGGAACAGCACTAAG 399856
QY 967 GTAGTATTGCAAAAGAAATATCTATCTATCTGATGTTGCTGGAATAATCAGAGATATT 1026
DB 399857 GCAACTGTGGATAAAGACCACACAACAAATCTGTTGAAGGTGCTGTTCTGCTGATGCTATT 399916
QY 1027 AAAGCTCGAGTTAAACAATTTGTCGCAAAATTCGAAGAAACAAGCTCAGATTATCATCGT 1086
DB 399917 TCTGATCGTGTAGCTATTATTAAAAAGCGCAAAATGAAAAACAATCTTGAATTTTGTAT 399976
QY 1087 GAAAACTTCAAGAACGCTTTGCAAACTCTGTTGGTGGAGTAGCTGTTTATCCATCTTGGA 1146
DB 399977 GAAAAATTACAAAGAACGTTCTTGCAAAATTAGCTGTGTGTTGCAAGTTGTTAAAGTTGT 400036
QY 1147 GCTGCTACTGAACTGAAATGAAAGAGAGAGGATCGGTAGAAAGATGCTCTTAAATGCA 1206

Db 856 ATCGTCGTAAGCTATGCTTGAAGACATTGCTATCTTGACAGGTGGTACAGTGATTACAG 915
Qy 908 AAGATCGTGTGTAAAGCTTGAAGATGTAAGCTTGTCTTTTAGAAGACAGCTAAACGTG 967
Db 916 AGGATCTAGGACTTGAATTAAGAAGATGCTACAATGACAGCCCTTGGACAGGCTGCTAAGA 975
Qy 968 TAGCTTATGCAAGAAGAAATACTACTATCTGATGCTGTGGAATATCAGAAATCAGAAATATTA 1027
Db 976 TTACAGTTGATAAGATAGCAGCAGTAACTTGTGAAGTTTCAAGAGTTTCAAGAGCTATTG 1035
Qy 1028 AAGCTCGAGTTAAACAAATTCGTGCAAAATGAAGAAACAGCTCAGATTATGATCGTG 1087
Db 1036 CTAACCGTATTGCACCTGATTAAATCGCAATTAGAAACAACACTTCTGACTTTGACCGTG 1095
Qy 1088 AAAAATCTAAGAACGCTTCTGAAAATCTTGTGTGAGTAGCTGTTTATCCATGTTGGAG 1147
Db 1096 AAAAATCTAAGAACGCTTCTGAAAATCTTGTGTGAGTAGCTGTTTATCCATGTTGGAG 1155
Qy 1148 CTGCTACTGAACCTGAATGAAGAGAGAGGATCGTGTAGAGATGCTCTAAATGCAA 1207
Db 1156 CTCACACAGAGACGCTTTAAAGAAATGAACTTCGATTTGAGGATGCTCTAAATGCTA 1215
Qy 1208 CAAGAGCTGCGTTGAAGAGGTATTCTCCCTGCTGTGTGCTACTGCTTTTGTCCGCTCOA 1267
Db 1216 CAGTGCAGCCGTTGAAGAGGTATCGTGTGTGTGGAACAGCAGCTTATTACGGTTA 1275
Qy 1268 TTAAGTCCTTGATGATTAATTAACCTGCTGATGATGATGAACTTGTGCTGGAGTTAATATCA 1327
Db 1276 TTGAAAAAGTAGCAGCTCTTGAGCTTGAGGGCGATGATG---CTACTGGAGCTAACATTG 1332
Qy 1328 TCCGTCGTTCTTGAAGAGCCTTTAGCTCAAAATGCTGCAAAATGCTGGCTATGAAGTT 1387
Db 1333 TGCTTCGTCCTCTAGAAGAGCCTGTAGCTCAAAATGCTTAAATGCTGGGTACGAGGCT 1392
Qy 1388 CTATTGTTGTAGAAAAAGTTCTGCAACCAAAAGATGTTTGGATTTAATGCTGCATCAG 1447
Db 1393 CCGTAGTTATTGACAGTTGAAGACAGCCCTGCGGAACAGGATTTAATGCTGCAACAG 1452
Qy 1448 GAGAAATGAGACCTTATTAAAGCTGTGTGTCATTGATCCTTAAAGTTACAGTATTG 1507
Db 1453 GTGAGTGGTTGATATGATTAACACAGGAATCAATGACCCCTGTCAAAGTAAACAGGATCAG 1512
Qy 1508 CATTACAAATGCAGCATCAGTAGCCTCTTACTTCTAACTACAGATGCGCTATTGCTG 1567
Db 1513 CGCTTCAAAATGACAGCTTCTGTAGTCTATTTTTGACACAGACGAGTTGTTGCTA 1572
Qy 1568 AAAAACCCAGAAC 1580
Db 1573 ATAACTCTGAAC 1585

Search completed: January 29, 2004, 13:35:33
Job time : 372 secs

This Page Blank (uspio)

Result No.	Query		DB	ID	Description
	Score	Match length			
1	1866	62.9	2	JC2562	chaperonin groELx
2	1864	62.8	545	2	groEL2 chaperonin
3	1863	62.8	545	2	heat shock protein
4	1862.5	62.8	547	2	60KDa chaperonin X
5	1852.5	62.4	546	2	chaperonin groEL -
6	1845	62.2	550	2	60K heat shock pro
7	1844.5	62.1	546	2	heat shock protein
8	1843.5	62.1	547	2	Cpn60 protein (Gro
9	1843.5	62.1	552	2	chaperonin 60 - Co
10	1841	62.0	542	2	probable heat shock
11	1840.5	62.0	544	2	60 KDa chaperonin
12	1840.5	62.0	544	2	60K chaperonin (pro
13	1828.5	61.6	544	2	groEL protein - Ag
14	1825.5	61.5	547	2	GroEL protein PA43

QY 1 ATGGCTTTCTAAAGAAATCTTTTGTATGCTAAAGCCCGTGAAGAACTTTTCACGAGGTGA 60
Db 1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuArgGlyVal 20
QY 61 GATAAACTTGTCAATGCTGTAAAGTAAACACATCTGGACCTTAAAGCCCGTGAATGCTTATT 120
Db 21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValle 40
QY 121 GAAAGCTTTTGGTTCCTCCAGTATTACAAAGATGGTGTATCTCTGTCGAAAGAAAT 180
Db 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluile 60
QY 181 GAACCTTGAAGATCAAGTTGAAATATGGCGCTCAATGGTTAAAGATAGCTCCCAAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80
QY 241 ACTGAGCGATATCTGCTGATGAACATACAAACAGCAACAGCTCTTGCACAAAGCTATTAT 300
Db 81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaVal 100
QY 301 CGTGAAGGTGTAACACTTGTAGCAGCTGGTGTATCTATGCGCAATTAACCTGGCATA 360
Db 101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyile 120
QY 361 GATAAAGCTGTTGCTGCTTACTTAAGAACTAAGCGACATTACAAAGCCCTACTCGTGAC 420
Db 121 AspLeuAlaValAlaGluValValLysAspLeuAlaLysAlaLysLysIleAsnThr 140
QY 421 CAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAACCTCTGATACAACTAGGTAAAT 480
Db 141 SerAspGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluLysGlnIleGlyLeu 160
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGGAAGAGGAGGTGTATTCACAGTTGAGGAAGCT 540
Db 161 AspIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180
QY 541 AAAGCTTCTGAACATACATTAGATGCTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
Db 181 LysThrAlaGluThrGluLeuGluValValGlyMetGlnPheAspArgGlyThrLeu 200
QY 601 TCTCCATCTTGTAACTCTGAGAAATGTTTGTGAACCTGTGAACCTTATATC 660
Db 201 SerProLysPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaPheile 220
QY 661 CTTTGTATGAGAAAGATTTACTAGCATGAAGACATGCTTACCAATCTTTAGAACAGTT 720
Db 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAACCGTCCACTCCTTATTATTCCTGAAGACCTAGAGGTGAGCACTGCA 780
Db 241 ValGlnThrGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATCTGTAGTCAATAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTCGGT 840
Db 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGCTGAAGCTATCTCTGAAGATATGCTATCTTACTGTGAGGAGAGCA 900
Db 281 PheGlyAspArgGlyLysAlaMetLeuGluAspIleAlaIleLeuThrGlyThrVal 300
QY 901 ATATTTGAAGATCGTGGTATTAAGCTTGAATGTAACTGTCTCTTTAGGAACAGCT 960
Db 301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320
QY 961 AAACGTGTAGTTATTGACAAAGAAAATACTACTACTGTTGATGCTGCTGGAAGATCAGAA 1020
Db 321 LysLysValSerIleThrLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340
QY 1021 GATATTAAAGCTCGAGTTTAAACAAATTCGTGCACAAATTAAGAAACAAAGCTCAGATTAT 1080
Db 341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTGAAGAACTCAAGAACGCTCTTGCAGAAACCTTGTGGTGGAGTAGCTGTTATCCAT 1140

Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleArg 380
QY 1141 GTTGGAGCTGCTACTCAAACTGAATGAAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
Db 381 ValGlyGlyAlaThrGluValGluValLysGluLysAspArgIleAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCTCGGTGGTACTGCTTTGTC 1260
Db 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420
QY 1261 CGC---TCCATTAAAGTCTCTGATGATATTAAACCTGCTGATGATGATGATGCTGGA 1317
Db 421 ArgSerSerValLysIle-----ThrValLysGlyGluAsnAspGlnAspAlaGly 438
QY 1318 CTTAATATCATCGTCTGCTCTTGAAGAGCCCTTACGTCAAAATGCTGCAATGCTGGC 1377
Db 439 ValAsnIleValArgAlaLeuGlnSerProAlaArgGlnIleValGluAsnAlaGly 458
QY 1378 TATGAAGGTCTATTCTGTAGAAAAGTTGCGGAACCAAAA---GATGTTTGGATT 1434
Db 459 AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspPheGlyTyr 478
QY 1435 AATGCTGCATCAGAGATATGAAGACCTTATTATAAGCTGGTGTCTCATTTGATCTCAAAAA 1494
Db 479 AsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys 498
QY 1495 GTTACACGATTATGCTATCAAAATGACGATCAGTAGCTCTTCTTCTTAACTACAGAA 1554
Db 499 ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu 518
QY 1555 TGGCTATTCTCTCAAAAACCAAGAACCTTAAAGATATGCTATGCTGCGGTGGTATG 1614
Db 519 AlaMetIleAlaGluLeu-----ProLysLysAspAlaProAlaMetProGlyGlyMet 536
QY 1615 GGTGGTATGCTGTATGACGGTATG 1641
Db 537 GlyGlyMetGlyGlyMetAspMetMet 545
RESULT 3
JN0509
heat shock protein groEL (clone Rhz A) - Rhizobium meliloti
N;Alternate names: chaperonin groEL protein
C;Species: Rhizobium meliloti
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C;Accession: JN0509
R;Rusanganwa, E.; Gupta, R.S.
Gene 126, 67-75, 1993
A;Title: Cloning and characterization of multiple groEL chaperonin-encoding genes in Rhiz
A;Reference number: JN0509; MUID:93231539; PMID:80971179
A;Molecule type: DNA
A;Residues: 1-545 <RUS>
A;Cross-references: GB:M94192; NID:gi52233; PIDN:AAA26285.1; PID:gi52235
C;Comment: This protein plays a role in protein folding and in the extracellular transpor
C;Genetics:
A;Gene: groEL
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; molecular chaperone; stress-induced protein

Alignment Scores:
Pred. No.: 1,28e-94 Length: 545
Score: 1863.00 Matches: 387
Percent Similarity: 81.06% Conservative: 58
Best Local Similarity: 70.49% Mismatches: 98
Query Match: 62.77% Indels: 6
DB: 2 Gaps: 4

US-09-077-574A-1 (1-1647) x JN0509 (1-545)

QY 1 ATGCTTCTTAAAGAAATCTTTTGTATGCTTAAAGCCCGTGAAGAACTTTTCACGAGGTGA 60
Db 1 MetAlaAlaLysGluValLysPheGlyArgSerGlyArgGluLysMetLeuArgGlyVal 20

Qy	1141	GTTCGACCTGCTACTGAACCTGAATCAAGAGAGAAGCATCGTAGAAGATGCCTTA	1200
Db	381	VaIGlygYAlaThrGLuValGLuValGLuLysAspArgIleAspAspAlaLeu	400
Qy	1201	AATGCCAACAGAGCTGCGGTGTAAGAAGGTATTGTCCTCGTGGTGCTACTGCTTTGTC	1260
Db	401	AsnAlaThrArgAlaAlaValGinGluGlyIleValProGlyGlyValAlaLeuLeu	420
Qy	1261	CGC---TCCATTTAAAGTCCTTGATGATATAACCTGCTGATGATGATGAACCTGCTGGA	1317
Db	421	ArgSerSerValIlysIle-----ThrValLysGlyGluAsnAspGlnAspAlaGly	438
Qy	1318	CCTTAATATCATCCGTCGTCCTCTTGAAAGACCCTTTACGTCAAATGCTGCAAAATGCTGGC	1377
Db	439	ValAsnIleValArgAlaLeuGinSerProAlaArgGlnIleValGluAsnAlaGly	458
Qy	1378	TATGAAGGTCTATTGTTGTAGAAAAGTTCTGTGAACCAAAA--GATGTTTTGGATTT	1434
Db	459	AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspAspPheGlyTyr	478
Qy	1435	AATGCTGCATCAGAGAAATATGAAGACCTTATTAAAGCTGGTCTCATTTGATCTTAAAAA	1494
Db	479	AsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys	498
Qy	1495	GTTACAGCTATTGCATTACAATGTCAGCATAGTGGCTTCCTTACTTCTTACTACAGAA	1554
Db	499	ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu	518
Qy	1555	TGCGCTATTGCTCAAAAAACCAAGACCTTAAAAAGATATGCTATGCTGCGGTGGTATG	1614
Db	519	AlaMetIleAlaGluLeu-----ProLysLysAspAlaProAlaMetProGlyGlyMet	536
Qy	1615	GGTGGTATGGGTGGTATGACCGGTATG	1641
Db	537	GlyGlyMetGlyGlyMetAspMetMet	545

३

(cc)

000C-ues-

0007-For

for Nucleation

4
0
0
0
4
0
6
2

1
0
1

osa.

below

AF83425.1; GSF

M.; Alvarenc

craro, D.M.; C

A.J.S.

11

; Franco, M. C.

ALLAN, E. E.

ques, M. V.; Mival

and, P. C., Myan
and, P. C., Palmi

Montelli, R.V.

Jr., W.A.: da

S.; Vettore,

Score: 1862.50 Matches: 366
Percent Similarity: 81.39% Conservative: 80
Best Local Similarity: 66.79% Mismatches: 97
Query Match: 62.75% Indels: 5
DB: 2 Gaps: 2

US-09-077-574A-1 (1-1647) x P82783 (1-547)

```
QY 1 ATGGCTCTTAAGAAATCCCTTTTGTGCTAAAGCCCGTGAAGAAATCTTACAGAGGTGTA 60
DB 1 MetAlaAlaLysGluIlePheSerGluLysAlaArgSerArgMetValHisGlyVal 20
QY 61 GATAAATCTGCAATCTGTTAAAGTAAACACTTGGACCTAAAGCCGTAATGTCGTATT 120
DB 21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyHisValValLeu 40
QY 121 GAAAGCTCTTTGGTCCCGAGTATTACAAAGATGGTGTATCTGTTGTCGAAAGAAAT 180
DB 41 AspLysSerPheGlySerProIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACTTGAAGATAGTTGAAATATGGCGCTCAATGGTTAAAGTAGCTCCCAA 240
DB 61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80
QY 241 ACTAGCATATTCTGCTGATGCACTACCAACGACACAGCTCTCTGACACAGCTATTAT 300
DB 81 ThrAsnAspHisAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaLeuIle 100
QY 301 CGTGAAGGTGTAAGCTTTAGCAGCTGGTGGTAACTCTATGGCCATTAACAGCTGGCATA 360
DB 101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTCTGCTGCTGTTACTTAAAGAACTAAAGCAGACATTAAGAGCTCTCGTAC 420
DB 121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysIleSerLysProThrSerAsp 140
QY 421 CAAAGAAATAGCTCAAGTTGGACCATTTCTGCAACCTCTGATACCAACAAATAGTAAT 480
DB 141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160
QY 481 ATCATAGCTCAAGCTATGGCTTAAAGTTGGAAGAGGCTGTATCAAGCTTGAAGAGCT 540
DB 161 IleIleAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGly 180
QY 541 AAAGGCTCTTGAACACTACATTAGATGCTGTTGAAGAAATGAAGTTGACCGGTGCTACTC 600
DB 181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200
QY 601 TCTCCATCTTGTAACTAATCTCTGAGAAATGTTGTGCACTTGTATTAACCTTATATC 660
DB 201 SerProTyrPheIleAsnAsnGlnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220
QY 661 CTTTGTATGAGAAAGAGTACTAGTCAAGTAAAGACATGCTTACCAATCTTAGAACAAGTT 720
DB 221 LeuLeuHisAspLysLysIleSerValArgAspLeuLeuThrValLeuAspAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATCTGCTGAAGACGTAGAGGTGAAGCACTTGCA 780
DB 241 AlaLysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260
QY 781 ACATTGTAGTCAATAGCTCCGTTGGAGCACTCCAAGTTGTAGCCGTAAGAGCTCTGCT 840
DB 261 ThrLeuValValAsnAsnIleArgGlyIleIleLysValCysAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGCTGAAAGCTATGCTGAAGATATTGCTATCTCTTACTGGAGGAGAACA 900
DB 281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300
QY 901 ATATTGTAAGTCTGTTGTTAAAGCTTGAAGCTTCTGCTTCTTTAGGAACAGCT 960
DB 301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrThrSerHisLeuGlyLysAla 320
QY 961 AAACGTGTAGTATTGCAAGAAATACTACTATCTGTTGATGCTGCTGGAATAACAGAA 1020
```

```
DB 321 LysLysValArgValSerLysGluAsnThrThrIleAspGlyIleGlyAspAsnAsp 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTAAGAAACAAGCTCAGATTAT 1080
DB 341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTGAAGAACTTCAAGAACTGTTGCAAACTTTGTTGGTGGAGTAGCTGTTATCCAT 1140
DB 361 AspArgLysLysLeuGlnArgValAlaLysLeuAlaGlyValAlaLysLys 380
QY 1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAAGAGATCGTGTAGAAAGATGCTCTA 1200
DB 381 ValGlyAlaAlaThrGluValGluMetLysGluLysAlaArgValAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420
QY 1261 CGTCCATTAAAGCTCTTGTATGATATTAAACCTGCTGATGATGATGATGATGATGATGAT 1320
DB 421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyAlaAsnGluAspGlnThrHisGlyIle 440
QY 1321 AATATCATCGCTGCTCTCTTGAAGAGCCCTTTACGTCAAATTCGTGCAAAATGCTGCTAT 1380
DB 441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460
QY 1381 GAAGGTTCTATTGTTAGAAAGTTGCTGCAACAAAGATGCTGTTGCTGATTTAATGCT 1440
DB 461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnPheGlyTyrAsnAla 480
QY 1441 GCATCAGGAGATATGAAGACCTTATTAAAGCTGTTGATGATGATGATGATGATGATGATGAT 1500
DB 481 AlaThrGlyGluPheGlyAspMetValAsnLeuGlyIleLeuAspProThrLysValThr 500
QY 1501 CGTATTGCTATTCAAAATGACGATCAGTACGCTCTTACTTCTTAATCAGAAATGCTGCT 1560
DB 501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520
QY 1561 ATTGCTGAAACACCAAGAACCTTAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
DB 521 ValAlaGlu-----AlaProLysLysAspGluProThrProAlaAlaGlyGlyGly 538
QY 1612 ATCGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1635
DB 539 MetGlyGlyMetGlyGlyMetAsp 546
RESULT 5
B47073
Chaperonin GroEL - Chromatium vinosum
C:Species: Chromatium vinosum
C:Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text_change 26-Aug-1999
C:Accession: B47073
R:Perreyra, R.G.; Soncini, F.C.; Viale, A.M.
J. Bacteriol. 175, 1514-1523, 1993
A:Title: Cloning, characterization, and functional expression in Escherichia coli of char
A:Reference number: A47073; MUID:93186721; PMID:8444812
A:Accession: B47073
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-546 <PER>
A:Cross-references: GB:M99443; NID:q145007; PIDN:AAA23319.1; PID:q145009
A>Note: sequence extracted from NCBI backbone (NCBI:126968, NCBI:P:126970)
C:Superfamily: chaperonin groEL
C:Keywords: molecular chaperone
Alignment Scores:
Pred. No.: 4,79e-94 Length: 546
Score: 1852.50 Matches: 374
Percent Similarity: 81.02% Conservative: 70
Best Local Similarity: 68.25% Mismatches: 101
Query Match: 62.42% Indels: 3
Gaps: 2
DB:
```

US-09-077-574A-1 (1-1647) x B47073 (1-546)

QY 1 ATGGCTTCTAAAGAAATCTTTTGTATGCTAAAGCCGTGAAACCTTTTACAGAGTGTA 60
Db 1 MetSerAlaGlyAspValLysPheGlyGlyAspAlaArgValArgMetMetGluGlyVal 20
QY 61 GATAAACTTGCCTCTTAAAGTAACACACTTGAACCTTAAAGCCGTGATGCTGTTATT 120
Db 21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGTCTTTTGTCTCCAGTATTATCAAAAGATGGTGTATCTGTGCAAAAGAAAT 180
Db 41 GluLysSerPheGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACTTGAAGATAAGTTTGAATAATATGGCGCTCAAAATGGTTAAAGAACTAGTCCCAA 240
Db 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTGCTGTGATGGAACACTACAACAGCAACAGTCTCTTGCACAAAGCTATTAT 300
Db 81 ThrSerAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaMetVal 100
QY 301 CCGTGAAGTGTAAACCTTGTAGCAGTGTGTGTATCTATGCGCCATTAAGCGTCATA 360
Db 101 ArgGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyMet 120
QY 361 GATAAGCTGTGCTCTTACTAAAGAACTAAGCGACATTAACAAGCTACTCGTGAC 420
Db 121 AspLysAlaValGluAlaThrGluGluLysLysLeuSerLysProCysProArg 140
QY 421 CAAAAGAAATAGCTCAAGTGTGAACCACTTCTGCAAACTCTGATACAACTAGTAAT 480
Db 141 ProMetAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspSerIleGlyThr 160
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTGTGAAGAGGAGTGTATCACAGTTCAGGAAGCT 540
Db 161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180
QY 541 AAAGCTCTTGAACATACATTAGATGTTGTAAGGAAGTAAGTCTTCAAGCTGCTACCTC 600
Db 181 ThrSerLeuGlnAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyThrLeu 200
QY 601 TCTCATACTTTGTAACTAATCTCGAGAAATGGTTGTGAACCTTGAATACCTTATATC 660
Db 201 SerProLysPheIleAsnAsnGlnSerGlnSerAlaGluLeuAspAlaProLysIle 220
QY 661 CTTTGTATGAGAAAGATTTACTAGCATGAAGACATGCTACCAATCTTAGAAACAGTT 720
Db 221 LeuLeuLysAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlyVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATTGCTGAAGAGTGAAGAGTGAAGCACTTGA 780
Db 241 AlalysAlaGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACACCTGTAGTCAATAAGCTCCGTGGACACTCCAAGTTGTAGCCGTAAAAGCTCCTGCT 840
Db 261 ThrLeuValValAsnThrIleArgGlyIleValLysValCysAlaValLysAlaProGly 280
QY 841 TTTGTGTGAACCGCTAAAGCTATGCTTGAAGATTTGCTATCTCTTACTTGGAGGAGCA 900
Db 281 PheGlyAspArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyAlaThrVal 300
QY 901 ATATTGTAAGATCGTGGTATAAGCTTGAATAATGTAAGCTTGTCTTCTTGAACAGCT 960
Db 301 IleSerGluGluValGlyLysSerLeuGluLysAlaThrLeuThrAspLeuGlyThrAla 320
QY 961 AAACGTGTAGTATTACAAAGAAATACTACTATCTGTTGATGGTGTGCGAAATCAGAA 1020
Db 321 LysArgValGlnValGlyLysAspGluThrThrIleIleAspGlySerGlySerGluIle 340
QY 1021 GATATTAAAGCTCGAGTTAAACAATTTGTGCAAAATTTGAAGAAACAGCTCAGATTAT 1080

Db 341 AspIleLysAlaArgCysGluGlnIleArgAlaGlnValGluGluThrSerSerAspTyr 360
QY 1081 GATCGTGAATAAACTTCAAGAACGCTCTTGTGCAAAACTTGTGTGGAGTAGCTGTATCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTGGAGCTGCTACTCAAACTGAAATGAAAGAGAAAGGATCGTGTAGAAGATGCTCTA 1200
Db 381 ValGlyAlaAlaThrGluIleGluMetLysGluLysAlaArgValGluAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGCTTGAAGAGTATTGTCCCTGGTGGTGTACTGCTTTTGTGTC 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuVal 420
QY 1261 CGCTCCATTAAAGTCTCTGATGATTAATAACCTGCTGTATGATGATGATGATGATGAT 1320
Db 421 ArgAlaIleAlaAlaValLysAspLeuLysGlyAlaAsnHisAspGlnAspValGlyIle 440
QY 1321 AATATCATCCGTCTCTTGAAGAGCTTTCAGTCAAAATGCTGCAAAATGCTGCTAT 1380
Db 441 AlaIleAlaArgAlaMetGluGluProLeuArgGlnIleValAlaAsnAlaGlyGlu 460
QY 1381 GAAGGTCTATTGTTGTAGAAAAGTTCGTGAACCAAAAGATGGTTTGGATTAAATGCT 1440
Db 461 GluProSerValIleLeuHisLysValAlaGluGlyThrGlyAsnPheGlyThrAsnAla 480
QY 1441 GCATCAGGAGATATGAACACTTATAAGCTGTGTGCTGATGATGATGATGATGATGAT 1500
Db 481 AlaAsnGlyGlyThrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThr 500
QY 1501 CCTATTGCAATCAAAATGCAGATCAGTAGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
Db 501 ArgSerAlaLeuGlnAsnSerCysSerValAlaGlyLeuMetIleThrGluAlaMet 520
QY 1561 ATGCTGAAAAACCAAGACCTTAAAAAGATATGCTCT---ATGCTGGCGGTGTATGGGT 1617
Db 521 IleAlaAsp-----GluProLysAspAlaProAlaMetProGlyGlyGlyMetGly 538
QY 1618 GGTATGGTGGTGTATGACCGGTATG 1641
Db 539 AspMetGlyGlyMetGlyMetMet 546

RESULT 6

A41468 60K heat shock protein htpB - Legionella pneumophila

C:Species: Legionella pneumophila

C>Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text_change 15-Oct-1999

C:Accession: A41468; A41472

R:Hoffman, P.S.; Houston, L.; Butler, C.A.

Infect. Immun. 58, 3380-3387, 1990

A:Title: Legionella pneumophila htpAB heat shock operon: nucleotide sequence and express

A:Reference number: A41468; MUID:90382960; PMID:2205580

A:Accession: A41468

A:Molecule type: DNA

A:Residues: 1-550 <HOF>

A:Cross-references: GB:M31918

R:Samson, J.S.; O'Connor, S.P.; Holloway, B.P.; Plikaytis, B.B.; Carlone, G.M.; Mayer, J.

Infect. Immun. 58, 3154-3157, 1990

A:Title: Nucleotide sequence of htpB, the Legionella pneumophila gene encoding the 58-kD

A:Reference number: A41472; MUID:90354095; PMID:2117582

A:Accession: A41472

A:Molecule type: DNA

A:Residues: 3-188, 'S', 190-195, 'LIAVH', 202-347, 'T', 349-550 <SAM>

A:Cross-references: GB:M91673; GB:M35149; NID:G149689; PIDN:AAA25298.1; PID:G149690

C:Genetics:

A:Gene: htpB

C:Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.: 1.23e-93 Length: 550

Score: 1845.00 Matches: 367

Percent Similarity: 82.02% Conservative: 80

Best Local Similarity: 67.34% Mismatches: 96

Query Match: 62.16% Indels: 2 Gaps: 1
DB: 2 1
US-09-077-574A-1 (1-1647) x A41468 (1-550)
QY 7 TCTAAGAAATCTTTTTCATGCTAAAGCCGCTGAAAACTTTTCACGAGGTGTAGATAAA 66
DB 4 AAlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 23
QY 67 CTTGCAAAATGCTTTAAAGTAACACTTGGACCTAAAGCCGCTAAATCGCTTATTTGAAAAG 126
DB 24 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 43
QY 127 TCTTTTGGTTCCTCCAGTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAAGTAACATT 186
DB 44 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluLeuGluPhe 63
QY 187 GAAGATAAGTTTGAAATATGGCGCTCAATGTTTAAAGAACTAGTCCCAAACTAGC 246
DB 64 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 83
QY 247 GATATTGCTGGTGAATGCAATACACAGCAACAGTCTTGCACAAAGCTATTATTCGTGAA 306
DB 84 AspThrAlaGlyAspGlyThrThrAlaThrValLeuAlaArgSerLeuValGlu 103
QY 307 GGTGTAAGACTTGTAGACGTGTGTATCTTATCTGCTTAAAGCTGCTGATAGATAAA 366
DB 104 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLysAspLys 123
QY 367 GCGTTGTTGCTGCTTACTAAGAACTAAGCGCATTACAAAGCTACTCGTGACCAAAA 426
DB 124 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 143
QY 427 GAAATAGCTCAAGTTGAACCACTTCTGCAAACTCTGATACAACTATGATTAATATCATA 486
DB 144 AlaLeuAlaGlnValGlyThrLeuSerAlaAsnSerAspGluAlaLeuGlyAlaLeu 163
QY 487 GCTGAAGCTATGCTTAAAGTTGAAAGAGGTGTATCAGTTGAGGAGCTAAAGGT 546
DB 164 AlaGluAlaMetGluLysValGlyLysGluGlyValLeuThrValGluAspGlyAsnGly 183
QY 547 CTTGAACTACATTAGATGTTGTTGAGGATGAAGTTTACCGTGGCTACCTCTCTCCA 606
DB 184 LeuGluAsnGluLeuThrValValGluGlyMetGlnPheAspArgGlyThrLeuSerPro 203
QY 607 TACTTTGTAATCTCTGAGAAATGTTTGTGAACCTTGATACCCCTATATCTCTTGT 666
DB 204 TyrPheLeuAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheLeuLeu 223
QY 667 AATGAGAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTTGCTAAA 726
DB 224 ValAspLysLysValSerSerLeuArgGluMetLeuSerValLeuGluGlyValAlaLys 243
QY 727 GTAACCGTCCACTCTTATTGCTGAAGACGTAGAAGTGAAGCTTGAACACTTGAACACT 786
DB 244 SerGlyArgProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 263
QY 787 GTAGTCAATAAGCTCCCTGGAGCACTCCCAAGTTGTAGCCGTAAAGCTCTCTGTTTGGT 846
DB 264 ValValAsnAsnMetArgGlyLeuValLysValCysAlaValLysAlaProGlyPheGly 283
QY 847 GAACCGCTAAAGCTATGCTTGAAGATATCTGTTATCTTCTGAGGAGAGCAATATTT 906
DB 284 AspArgArgLysAlaMetLeuGluAspLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 303
QY 907 GAAGATCGTGTATAAGCTTGAATGTAAGCTTGTCTTCTTAGGAAACAGCTAAACGT 966
DB 304 GluGluLeuGlyLysSerLeuLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 323
QY 967 GTAGTATTGCAAGAAATACTACTATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1026
DB 324 IleValValThrLysLysGluAsnThrThrIleLeuAspGlyGluGlyLysAlaThrGluLeu 343

QY 1027 AAAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGATCGT 1086
DB 344 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluThrThrSerAspTyrAspArg 363
QY 1087 GAAACATTTCAAGACGCTTTCGCAAACTTGTGTGGAGTAGTCTGTTATCCATGTTGGA 1146
DB 364 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValLysValGly 383
QY 1147 GCTGCTACTCAAACTCAAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
DB 384 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 403
QY 1207 ACAAGAGCTCGCGTTCAAGAAAGTATTGTCCCTGGTGGTGGTGGTGGTGGTGGTGGT 1266
DB 404 ThrArgAlaAlaValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 423
QY 1267 ATTAAGTCTCTGATGATATTAACCTGCTGATGATGATGATGATGATGATGATGATGAT 1326
DB 424 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 443
QY 1327 ATCCGTCGTTCTTCAAGAGAGCTTTTACGTCAAATTTGCTCAATGCTGCTGCTGATGAGGT 1386
DB 444 LeuArgAlaAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 463
QY 1387 TCTATTGTTGTAGAAAAGTTTCGTGAACCAAAAGATGTTTTCGATTAAATGCTGCATCA 1446
DB 464 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 483
QY 1447 GGAGAATATGAACCTTATTAAAGCTGGTGTGCTCATGATCTAAAGTAAAAAGTTACAGTATT 1506
DB 484 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 503
QY 1507 GCATTCAAAATGCAGATCAGTACCTCTTCTTAACTACAGATCGCTGCTGCTGCTGCT 1566
DB 504 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 523
QY 1567 GAAAAACCAACCTAAAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1626
DB 524 AspLeu-----ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 541
QY 1627 GGTATGACCGTATG 1641
DB 542 GlyMetGlyGlyMet 546
RESULT 7
S65596
heat shock protein 60 - Rhizobium leguminosarum
N;Alternate names: chaperonin 60
C;Species: Rhizobium leguminosarum
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C;Accession: S65596
R;Wallington, E.J.; Lund, P.A.
Microbiology 140, 113-122, 1994
A;Title: Rhizobium leguminosarum contains multiple chaperonin (cpn60) genes.
A;Reference number: S65596; MUID: 94214663; PMID: 7909257
A;Accession: S65596
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-546 <WAL>
A;Cross-references: EMBL:L20775; NID:g387873; PIDN:AAA26246.1; PID:g387874
C;Genetics:
A;Gene: cpn60
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; molecular chaperone
Alignment Scores:
Pred No.: 1.31e-93 Length: 546
Score: 1844.50 Matches: 384
Percent Similarity: 80.91% Conservative: 61
Best Local Similarity: 69.82% Mismatches: 98
Query Match: 62.15% Indels: 7
DB: 2 Gaps: 5

US-09-077-574A-1 (1-1647) x S6596 (1-546)

```
QY 1 ATGGCTTCTAAAGAAATCTTTTGTATGCTTAAAGCCCGTGAAGAACTTTCAACAGGTGA 60
Db |||||
1 MetAlaSerLysGluLeuLysPheGlyArgThrGlyArgGluLysMetLeuArgGlyVal 20
QY 61 GATAAATCTGTAATCTGTTTAAAGTAAACACTTGGACCTTAAAGCCGCTAATGCTGTTAT 120
Db |||||
21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValIleIle 40
QY 121 GAAAGTCTTTGGTCCCGCTATTAACAAAAGATGGTGTATCTCTTGGCAAGAAAT 180
Db |||||
41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACTTGAAGATAAGTTTCAAAATATGGCGCTCAAAATGCTTAAAGAGTAGCTCCCAA 240
Db |||||
61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80
QY 241 ACTAGGATATTGCTGGTGAAGAACTAACAAGCAACAGTCTCTGCACAAGCTATTAT 300
Db |||||
81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CTGTAAGGTGTAATCTTACAGCTGCTGCTAATCTTATGCTTATGCTTAAAGCTGCTA 360
Db |||||
101 ArgGluGlyAsnLysAlaValAlaAlaGlyMetAsnProMetAspLysArgGlyIle 120
QY 361 GATAAGCTGTGTTGCTTACTTAAAGAACTAAGCGCATTACAAAGCTTACTCTGTGAC 420
Db |||||
121 AspLeuAlaValAlaAspValValLysAspLeuGlnAlaLysLysIleSerThr 140
QY 421 CAAAGAAATAGTCAAGTGGACCACTTTCGAAACTCTGCATACAAATAGTAAAT 480
Db |||||
141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyAspLysGlnValGlyLeu 160
QY 481 ATCATAGCTGAGCTATGCTTAAAGTGGAAAGAGGAGTGTATACAGTTGAGAAGCT 540
Db |||||
161 AspiIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluAla 180
QY 541 AAAGTCTTTGAACTACATTTAGATGTGTTGAAGAACTGAAGTTCACCGTGGCTACCTC 600
Db |||||
181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyThrLeu 200
QY 601 TCTCATATCTTGTAACTTCTGAGAAATGGTTTGTGAACCTTGAACCTTATATC 660
Db |||||
201 SerProTyrPheValThrAsnProGluLysMetIleAlaAspLeuGluAspValPheIle 220
QY 661 CTTTGTATGAGAAAGATTACTAGCATGAAGCATGCTACCAATCTTAGAACCAAGTT 720
Db |||||
221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnSerMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATTGCTGAAGAGCTGAAGGTTGAAGCTTCA 780
Db |||||
241 ValGlnThrGlyLysProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTGTAGTCAATAAGCTCGTGGAGCATCTCAAGTTGTAGCGTAAAGCTCCTCGT 840
Db |||||
261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280
QY 841 TTTGTTGAACCGCTGAAGCTATGCTTCAAGATATTGCTATCTCTTCTTAAAGCAAGCA 900
Db |||||
281 PheGlyAspArgLysArgMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300
QY 901 ATATTTGAAGATCGTGGTATAAAGCTTGAAGATGTAAGCTTCTCTTAAAGCAAGCT 960
Db |||||
301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320
QY 961 AAACGTGTAGTATTGACAAAGAAATCTACTATCTGTTGATGGTGTGGGAAATTCAGAA 1020
Db |||||
321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlySerGlyAlaLysThr 340
QY 1021 GATATTAAGCTCGAGTTTAAACAAATTCGTCACAAATTTGAAGAAACAGCTCAGATTAT 1080
Db |||||
341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360
```

```
QY 1081 GATCGTGAAGAACTTCAAGAACGCTCTTGCCTCAAACTTGTGCTGAGTAGCTGTATCCAT 1140
Db |||||
361 AspArgLysLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleArg 380
QY 1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAAGGATCGCTGAGAAGATGCTCTA 1200
Db |||||
381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGTATTGTCCTCGTGGTGGTACTGCTGTTGTC 1260
Db |||||
401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyIleAlaLeuAla 420
QY 1261 CGCTCCATTAAAGTCTTCATGATATTAACCTGCTGATGATGATGATGATGATGATGAT 1320
Db |||||
421 ArgSerSerThrLysIle---ThrValLysGlyAlaAsnAspAspGlnGluAlaGlyIle 439
QY 1321 AATATCATCGCTGCTTCTTGAAGAGCCTTTACGTCACAAATGCTGCAAAATGCTGCTAT 1380
Db |||||
440 AsnIleValArgArgAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459
QY 1381 GAAGGTTCTATTGTTGTAGAAAAGTTCTGTGAACCA---AAAGATGGTTTGGATTAT 1437
Db |||||
460 GluAlaSerIleValValGlyLysValLeuAspLysAsnGluAspAsnPheGlyTyrAsn 479
QY 1438 GCTGCATCAGGAGAAATATGAAGACCTTATTAAAGCTGCTGTCTCATGATCCTAAAAAGTT 1497
Db |||||
480 AlaGlnThrSerGluTyrGlyAspMetIleAlaMetGlyIleValAspProLeuLysVal 499
QY 1498 ACACGTATTGCTTCAAAATCGAGCATCAGTAGCTCTTACTTCTTAACTACAGAAATGC 1557
Db |||||
500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 519
QY 1558 GCTATTGCTGAAAAACCCAGAACCTAAAAAGATATGCTCT---ATGCTCGCGGTGGT 1611
Db |||||
520 MetIleAlaGluLeu---ProLysLysAspAlaProAlaGlyMetPro---GlyGly 536
QY 1612 ATGGGTGTATGGTGGTGTATGGACGGTATG 1641
Db |||||
537 MetGlyGlyMetGlyGlyMetAspMetMet 546
```

RESULT 8

I40331

Cpn60 protein (GroEL) - Bordetella pertussis

C.Species: Bordetella pertussis

C.Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text_change 26-Aug-1999

C.Accession: I40331

R.Fernandez, R.C.; Weiss, A.A.

Gene 158, 151-152, 1995

A.Title: Cloning and sequencing of the Bordetella pertussis cpn10/cpn60 (groESL) homolog

A.Reference number: I40330; MUID:95309719; PMID:7789805

A.Accession: I40331

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-547 <RES>

C.Cross-references: EMBL:U12277; NID:G968918; PIDN:AAA74967.1; PID:G968920

C.Genetics:

A.Gene: cpn60

C.Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	1,49e-93	Length:	547
Score:	1843.50	Matches:	370
Percent Similarity:	80.87%	Conservative:	74
Best Local Similarity:	67.40%	Mismatches:	98
Query Match:	62.11%	Indels:	7
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x I40331 (1-547)

QY 1 ATGGCTTCTAAAGAAATCTTTTGTATGCTTAAAGCCCGTGAAGAACTTTCAACAGGTGA 60

Db |||||

1 MetAlaLysGlnValLeuPheAlaAspGluAlaArgValArgIleValArgGlyVal 20

```
QY 61 GATAAAGCTTCAATGCTGTGTAAGTAACACTTGGACCTTAAGCCGTAATGCTGTTATT 120
Db 21 AsnValLeuAlaAsnAlaValLysThrThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGGCTTTTGGTTCCTCCAGTATTACAAAAGATGCTGTAATCTGTGTCGAAAGAAATT 180
Db 41 GluArgSerPheGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluLe 60
QY 181 GAATCTGAAGATGAAGTTGAAATATATGGCGCTCAATGTTTAAAGAGTAGTCCCAAA 240
Db 61 GluLeuLysAspLysPheGluAsnLeuGlyAlaGlnLeuValLysAspValAlaSerLys 80
QY 241 ACTAGCCATATTGCTGTGTGATGAGTCACTCAACAGCACAGTCTCTGCACAGCTATTAT 300
Db 81 ThrSerAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaValVal 100
QY 301 CGTGAAGGTGTAAGCTTGTAGCAGCTGGTGTGTAATCTTATGGCCATTAAACGTCGCATA 360
Db 101 GlnGluGlyLeuLysTyrValAlaAlaGlyPheAsnProLysLeuLysArgGlyLe 120
QY 361 GATAAAGCTGTGTGCTGTGTACTTAAGAACTTAAGCGACATTACAAAGCTTACTCGTGAC 420
Db 121 AspLysAlaValAlaAlaValGluGluLeuLysLysLeuSerLysProValThrThr 140
QY 421 CAAAAGAAATAGCTCAAGTTGGAACTTCTCAAACTCTGATACAACTAGTAATAGTAAAT 480
Db 141 SerLysGluLeuAlaGlnValGlySerIleSerAlaAsnSerAspAlaSerIleGlyGln 160
QY 481 ATCATAGCTGAAGCTATGGCTTAAAGTTGAAAGGAGGTGTATCACAGTTGAGGAAGCT 540
Db 161 IleIleAlaAspAlaMetAspLysValGlyGluGlyValIleThrValGluAspGly 180
QY 541 AAAGGCTCTGAACACTACATTAGATGTGTTGAAGGAATGAAGTTGACCGGTGCTACTC 600
Db 181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATCTTGTAACTAATCTCGAAGAAATGTTTGTGACTGTATGATACCTTATATC 660
Db 201 SerProTyrPheIleAsnSerProGluLysGlnValAlaAlaLeuAspAspProTyrVal 220
QY 661 CTTTGTATGAGAAAGATGACTAGCATGAAAGACATGCTTACCAATCTTAGAACAAGTT 720
Db 221 LeuIleTyrAspLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
QY 721 GCTAAAGTAACCGTCCACTCTTATTATCTCAAGACGTPAGAAGTGTAAGCACTTGCA 780
Db 241 AlaLysSerSerArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATCTGTAGTCAATAGCTCCGTGGAGCCTCCAAAGTTGTAGCGTAAAGCTCCTGCT 840
Db 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrThrAlaValLysAlaProGly 280
QY 841 TTTGCTGAACCGCTTAAAGCTATGCTGAAGATATTGCTATCTTACTGAGAGGAAGCA 900
Db 281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300
QY 901 ATATTGTAAGATCGTGTGTPATAAGCTTGAATGTAAGCTTGTCTTCTTTAGGAAACAGCT 960
Db 301 IleSerGluGluThrGlyMetSerLeuGluLysAlaThrLeuGlnAspLeuGlyGlnAla 320
QY 961 AAAGCTGTAGTTATTGACAAAGAAATACTACTACTATCTGTGATGCTGCTGGAATACGAA 1020
Db 321 LysArgIleGluValAlaLysGluAsnThrThrIleIleAspGlyAlaGlyAspGlyLys 340
QY 1021 GATATTAAAGCTCAGTTAAACAAATCGTCACAAATTAAGAAACAAAGCTTCAGATTAT 1080
Db 341 SerIleGluAlaArgValLysGlnIleArgAlaGlnIleGluGluAlaThrSerAspLys 360
QY 1081 GATCGTGAAGAACTTCAAGAACGCTTGTGCAAAACCTTGTGTGGAGTAGTACTGTTATCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleArg 380
```

```
1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAAGGATCGTGTAGAGATGCTCTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
QY 1201 AATGCAACAAGAGCTCGGTTGAAGAGTATTGTTCCCTCGTGGTGTACTGCTTTTGTCT 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuLeu 420
QY 1261 CGCTCCATTAAAGTCTTGTGATGATATTAAACCTGCTGATGATGATGAACCTTGTGGAATT 1320
Db 421 ArgAlaLysGlnAlaIleThrGlyLeuLysGlyAspThrAlaAspGlnAsnAlaGlyIle 440
QY 1321 AATATCATCTCGTCTTCTTGAAGAGCTTAACTGCTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db 441 LysLeuIleLeuArgAlaValGluGluProLeuArgThrIleValThrAsnAlaGlyAsp 460
QY 1381 GAAGGTTCTATTGTTGAGAAAAGTTCGTGAACCAAGATGCTTTTGGATTTAATGCT 1440
Db 461 GluAlaSerValValValAsnThrValLeuAsnGlyLysGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGGAAATATGAAGACCTTATTAAAGCTGCTGCTCATTCATCTTAACTAAGATTTACA 1500
Db 481 AlaThrGlyGluTyrGlyAspLeuValGluGlnGlyValLeuAspProThrLysValThr 500
QY 1501 CGTATTGCATTACAAAATGAGATCAGTACGCTCTTCTTCTTAACTAAGATTTAAGATGCT 1560
Db 501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuLeuThrAlaGluAlaAla 520
QY 1561 ATTGCTGAA-----AAACAGAACCTTAAAGATATGCTTATGCTGCTGCGGT 1608
Db 521 ValValGluLeuMetGluAsnLysProAlaAlaAlaProAlaMetPro-----Gly 537
QY 1609 GGTATGGCTGGTATGGGTGATGGAC 1635
Db 538 GlyMetGlyGlyMetGlyGlyMetAsp 546

RESULT 9
S39765
Chaperonin 60 - Coxiiella burnetii
C:Species: Coxiiella burnetii
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 28-May-1999
C:Accession: S39765
R:Vodkin, M.H.; Williams, J.C.
J: Bacteriol. 170: 1227-1234, 1988
A:Title: A heat shock operon in Coxiiella burnetii produces a major antigen homologous to
A:Reference number: S39764; MUID:88139182; PMID:3343219
A:Accession: S39765
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <VOD>
A:Cross-references: GB:M20482; NID:g144996; PIDN:AAA23309.1; PID:g144998
C:Superfamily: chaperonin groEL

Alignment Scores:
Pred. No.: 1,49e-93 Length: 552
Score: 1843.50 Matches: 365
Percent Similarity: 81.42% Conservative: 82
Best Local Similarity: 66.48% Mismatches: 99
Query Match: 62.11% Indels: 3
DB: 2 Gaps: 3

US-09-077-574A-1 (1-1647) x S39765 (1-552)
QY 1 ATGCTCTTAAGAAATCCCTTTTGTATGCTTAAGCCCTGAAACCTTTACAGAGTCTA 60
Db 1 MetAlaLysValLeuLysPheSerHisGluValLeuHisAlaMetSerArgGlyVal 20
QY 61 GATAAAGCTTCAATGCTGTAAAGTAACTCTGACCTAAGCGCTAAGCGCTAAGTCTGATT 120
Db 21 GluValLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGCTCTTTGTTTCCCAAGTTATTACAAAGATGCTGTATCTGTTGCTGCAAAAGAAATT 180
```

```

Alignment Scores:
Pred. No.:          2.04e-93      Length:          542
Score:             1841.00       Matches:         373
Percent Similarity: 81.32%      Conservative:    71
Best Local Similarity: 68.32%   Mismatches:     96
Query Match:       62.03%      Indels:         6
DB:                2           Gaps:           2

US-09-077-574A-1 (1-1647) x F95967 (1-542)

QY   1  ATGGCTTCTAAAGAAATCCTTTTGTGCTAAAGCCCGTGAAAAACTTTTCACGAGGTGA 60
      :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   1  MetAlaAlaLysGluVallyspheGlnThrAspAlaArgGluArgMetLeuArgGlyVal 20

```

Qy	61	GATAAAC	TTGCAAT	GTGTTAA	AGTAACA	CTTCGAC	CTAAAG	CCCGTAAT	GTCTGTTAT	120
Db	21	AspValLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValle	40							
Qy	121	GAAAGCT	TTTGGT	CCCCAG	TTATTACA	AAAGAT	GGTGAT	CTGTG	TGCAAAAGAAAT	180
Db	41	AspLysSerPheGlyAlaProArgGlyThrLysAspGlyValSerValAlaLysGluLe	60							
Qy	181	GAACCT	GAAAGTA	AGTTG	AAAAAT	TGGCCCT	CAAAAT	TGGTT	TAAAGAGTAGCTCCCAAA	240
Db	61	GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerArg	80							
Qy	241	ACTAGC	GAATAT	TGCTGT	GATGGA	CACTACA	CAGCA	CAGTCT	CTTGCAACAGCTATTAT	300
Db	81	ThrAsnAspLeuAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaVal	100							
Qy	301	CGTGA	AGGTG	TAAAC	TTGTAG	CAGCT	GGTGT	ATCCT	TATGGCCATTTAAACGTTGGGATA	360
Db	101	ArgGluGlyAlaLysAlaValAlaSerGlyMetAsnProMetAspLeuLysArgGlyLe	120							
Qy	361	GATAA	AGCTGT	TGTTG	CTTACT	TAAAGAA	CTAAAG	CGACAT	TACAAAGCCCTACTCGTGAC	420
Db	121	AspLeuAlaValAspAlaValValLysGluLeuLysAsnAsnAlaArgLysIleSerLys	140							
Qy	421	CAAAA	GAAATAG	CTCA	AGTTG	GAAAC	CAATTTCT	GCAAACT	CTGTATACACAATAGGTAA	480
Db	141	AsnSerGluIleAlaGlnValGlyThrIleSerAlaAsnGlyAspThrGluIleGlyArg	160							
Qy	481	ATCAT	AGCTGA	AGCTAT	GGCTAA	AGTTG	GAAAGG	AGGTGT	TATCACAGATTTGAGGAGCT	540
Db	161	TyrLeuAlaGluAlaMetGluLysValGlyAsnGluGlyValIleThrValGluGluAla	180							
Qy	541	AAAGG	CTTGAA	ACTAC	TATAG	ATGTGG	TTCAGG	AAATGA	AGTTTGACCGTGGCTACCTC	600
Db	181	LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu	200							
Qy	601	TCTCC	ACTTTT	GTAACT	TAATCT	CGAGAAA	ATGTTT	GTGAAC	TTGTATPAACCCCTTATATC	660
Db	201	SerProTyrPheIleThrAsnGlnAspLysMetArgValGluLeuGluAspProTyrIle	220							
Qy	661	CTTTG	TATATG	AAAAAG	ATTACT	TAGCAT	GTAAGAC	ATGCTT	ACCAATCTTAGACAAGTT	720
Db	221	LeuIleHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal	240							
Qy	721	GCTAA	AGTAA	ACCGT	CCACT	CCCTT	TATTCT	CTCAAG	ACGTAGAGGTGAAGCACTTGCA	780
Db	241	ValGlnSerGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla	260							
Qy	781	ACACT	GTATG	TACAT	AGCTCC	GTGGAG	CACTCCA	AGTTGT	TAGCGGTAAAGCTCTGGT	840
Db	261	ThrLeuValValAsnLysLeuArgGlyLysValAlaAlaValLysAlaProGly	280							
Qy	841	TTTCG	TGNA	CCCGT	TAAGCT	ATCTT	GACAT	ATTCAT	CTCTCTCTCTAGGACACACT	900
Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal	300							
Qy	901	ATATT	TGAAG	ATCG	TGTATA	AAAGCT	CTCAAA	ATGTAA	CGTCTCTTCTTAGGACACACT	960
Db	301	ValSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla	320							
Qy	961	AAACG	TGTAG	TATTG	CAAAA	AAAAAT	ACTACT	ACTCT	GTGTTGATGTTGGTCTGGAATTCAGAA	1020
Db	321	LysLysValSerIleGluLysGluAsnThrThrIleIleAspGlyAlaGlySerLysAla	340							
Qy	1021	GATATT	AAAGCT	CGAGTT	TAAACAA	ATTG	CTGTG	CAAAAT	TGAAGAAAACAAGCTCAGATTAT	1080
Db	341	AspIleGluGlyArgThrAlaGlnIleArgAlaGlnIleGluGluThrThrSerAspTyr	360							
Qy	1081	GATCG	TGAAA	AACTT	CAAGA	CGTCT	CCAAA	CTTGT	TGTTGGTAGTAGCTGTATCCAT	1140
Db	361	AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg	380							
Qy	1141	GTTGG	AGCTG	CTACT	GAAAC	CTGA	ATTA	GAAAG	AGAGAAAGGATTCGTTGAGAAGATGCTCTA	1200

```

||||| :::||||| |||||:::|||||:::|||||:::|||||:::|||||:::|||||
381 ValGlySerThrGluValGluValGlyGlyLysAspArgValAspAlaLeu 400
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1201 AATGCAACAAGAGCTGGGTGTAAGAAGGATTGTCCTCGTGGTGACTCTGTTGTC 1260
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
401 HisAlaThrArgAlaAlaValGluGluGlyLeuProGlyGlyGlyValAlaLeu 420
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1261 CGTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGAACCTTGCTGACTT 1320
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
421 ArgAlaValLysAlaLeuAspGlyLeuLysThrAlaAsnAspGlnArgValGlyVal 440
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1321 AATATCATCGTGGTCTCTTGAAGACCTTTAGCTCAAAATTGCTGCAAAATGCTGGCTAT 1380
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
441 AspLeuValArgAlaIleGluAlaProValArgGlnIleAlaGluAsnAlaGlyAla 460
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1381 GAAGGTTCTATTGTTGAGAAAAAGTTGCTGAACCAAAAGAT--GGTTTTGGATTAAAT 1437
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
461 GluGlySerIleIleValGlyLysLeuArgGluLysThrGluPheSerTyrGlyTyrAsn 480
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1438 GCTGCATCAGGAGAAATGAAGACCTTTATTAAAGCTGGTGCAATTGATCTCTAACTACAGATGC 1497
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
481 AlaGlnThrAsnGluTyrGlyAspLeuTyrAlaMetGlyValIleAspProAlaLysVal 500
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1498 ACAGTATTGCATTACAAATGACGACCTAGTAGCCTCCCTACTTCTTAACTACAGATGC 1557
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla 520
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1558 GCTATTCTGCAAAACCAGAACCTTAAAGATATGCTATGCTGCGCGGTGGTATGGGT 1617
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
521 MetIleAlaGluLysProLysLysGluAlaAlaProAlaLeuProAlaGly----- 537
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1618 GGTATGGTGGTATGGAC 1635
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
538 -----GlyGlyMetAsp 541
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
RESULT 11
AD2660
60 KDA chaperonin [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD2660
F:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.;
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Goid-
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2660
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAU41698.1; PID:gl7739044; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Map position: circular chromosome
A:Gene: groEL
C:Superfamily: chaperonin groEL

Alignment Scores:
Pred. No.: 2.17e-93 Length: 544
Score: 1840.50 Matches: 376
Percent Similarity: 81.95% Conservative: 69
Best Local Similarity: 69.24% Mismatches: 95
Query Match: 62.01% Indels: 3
DB: 2 Gaps: 3

US-09-077-574A-1 (1-1647) x AD2660 (1-544)

Qy 1 ATGCTCTTAAAGAAATCCCTTTTGATGCTTAAGCCCGTGAAGAACTTTCACGAGGTGA 60
Db 1 MetAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuLysGlyVal 20

```


Db	21	AspIleuAlaAspAlaValLysValThrLeuGlyProLysGlyAsnValValIle	40
Qy	121	GAAGAAGCTTTTGGTCCCAAGTTATTACAAAAGATGGTGATCTGTGGTCAAAAAGAAATT	180
Db	41	AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle	60
Qy	181	GAACCTGAAGATAAGTTTGAAAATATATGGCGCTCAATGGTGTAAAGTAAGTAGTCCCAAA	240
Db	61	GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys	80
Qy	241	ACTAGCGATATTCGTGGTGAGGAACACTACACAGCAACAGTCCCTGCACAGCTATTAT	300
Db	81	ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal	100
Qy	301	CGTCAAGGTGTAAACCTTGTACAGCTGGTCTAATCTATGGCCATTAAACGTGGGATA	360
Db	101	ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle	120
Qy	361	GATAAAGCTGTGTGTGTCTTACTAAAGAACTAAAGCGACATTACAAAGCTTACTCGTGAC	420
Db	121	AspLeuAlaValAlaGluValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr	140
Qy	421	CAAAAAGAAATAGCTCAAGTTGGACCACTTTCGCAAACTCTGATACAACTAGGTAAAT	480
Db	141	SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu	160
Qy	481	ATCATAGCTGAAGCTATCGCTAAAGTTGCAAAAGGAGGTGTATCACAGTTGAGGAAGCT	540
Db	161	AspIleAlaGluAlaMetGluArgValGlyAsnGluGlyValIleThrValGluGluAla	180
Qy	541	AAAGGCTCTGAACCTACATTAGATGTGGTTCAAGGAATGAAGTTTGACCGTGGCTACCTC	600
Db	181	LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu	200
Qy	601	TCTCCATCTTTGTAACTAATCTCGAGAAATGTTTGTGAACCTGATACCCCTTATATC	660
Db	201	SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle	220
Qy	661	CTTTGTTAATGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTT	720
Db	221	LeuLeuHisGluLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal	240
Qy	721	GCTAAGTAAACCCGTCCTCTTATTATGCTGAAGACGTAGAGGTGAAGCACTTGCA	780
Db	241	ValGlnThrGlyLysProLeuValIleIleAlaGluAspValGluGlyGluAlaLeuAla	260
Qy	781	ACACTGTAGTCAATAAGCTCGTGAGACATCCCAAGTTGTAGCCGTAAAAGCTCTCGT	840
Db	261	ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly	280
Qy	841	TTTGTGTAACCGTAAAGCTATGCTCAAGATATTGCTATCTCTTCTTTTAGGAAACAAGT	900
Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal	300
Qy	901	ATATTGAAGATCGTGGTATAAAGCTTGAAATGTAAGCTTGTCTCTTTTAGGAAACAAGT	960
Db	301	IleSerGluAspLeuGlyLysLeuGluSerValThrLeuAspMetLeuGlyLysSer	320
Qy	961	AAACGTGTAGTTATTGCAAGAAATACTACTATCTGTTGATGGTCTGGGAAATACAGAA	1020
Db	321	LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyValAlaGlyGlnLysSer	340
Qy	1021	GATATTAAAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAGCTCACATTAT	1080
Db	341	AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr	360
Qy	1081	GATCGTGAAGAACTTCAAGAAAGCTCTTGCAAACTTGTGTGGTAGTAGCTGTTATCCAT	1140
Db	361	AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg	380
Qy	1141	GTTGAGTGTCTGAAACTGAAATGAAGAGAGAGGATCGTGTAGAGATGCTGCTTA	1200
Db	381	ValGluGlyLysThrGluValGluValLysGluLysLysAspArgIleAspAlaLeu	400

Qy	1201	AATGCAACAAGAGCTCGGTGTGAAGAAGGTATTGTCTCCTGGTGGTGTACTGCTCTTTGTC	1260
Db	401	AsnAlaThrArgAlaAlaValGlnGluGlyLeuValProGlyGlyGlyValAlaLeuLeu	420
Qy	1261	CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACCTTGCTGGACTT	1320
Db	421	ArgSerSerThrLysIle---ThrValLysGlyValAsnAspGlnGluAlaGlyIle	439
Qy	1321	AAATATCATCGTCTGCTTCTTGAAGAGCCCTTACGTCAAAATTCGTCAAAATTCGTGGCTAT	1380
Db	440	AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp	459
Qy	1381	GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCA---AAAGATGGTTTGGATTTAAT	1437
Db	460	GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn	479
Qy	1438	GCTGCATCAGAGAAATGAAGACCTTATTAAAGCTGGTGTTCATTGATCTCTAAAAAAAGTT	1497
Db	480	AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal	499
Qy	1498	ACAGTATTGCATTACAAATATGCGAGCATCAGTAGCCCTCTTCTTCTTAACCTACAGATGC	1557
Db	500	ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla	519
Qy	1558	GCTATTCTCGTAAAAACAGAACCTTAAAAAAGATATGCCT---ATGCTGGCGGGTGTATG	1614
Db	520	MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyGlyMet	539
Qy	1615	GGTGGTATG 1623	
Db	540	GlyGlyMet 542	
RESULT 13			
S23918			
groEL protein - Agrobacterium tumefaciens			
C/Species: Agrobacterium tumefaciens			
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999			
C/Accession: S23918			
R/Segal, G.; Ron, E. Z.			
submitted to the EMBL Data Library, August 1992			
A/Description: Cloning and sequencing of the groE operon of Agrobacterium tumefaciens			
A/Reference number: S23917			
A/Accession: S23918			
A/Status: preliminary			
A/Molecule type: DNA			
A/Residues: 1-544 <SRG>			
A/Cross-references: EMBL:X68263; NID:g1019913; PIDN:CAA48331.1; PID:g39095			
C/Superfamily: chaparrin groEL			

Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80
QY 241 ACTAGCGATATCTCGTGTGATGAACTACAAACAGCAACAGCTCTTCACCAAGCTATTAT 300
Db 81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CGTGAAGGTGTAACACTGTAGCAGCTGGTGGTAACTCTATGCGCTTAAACCTGGCATA 360
Db 101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLysIleArgGlyLeu 120
QY 361 GATAAAGCTGTGTTCTCTGTTTACTAAAGAACTAAGCGACATTAACAAGCCCTACTCGTAC 420
Db 121 AspLeuAlaValAlaGluValValLysAspLeuGlnAlaLysAlaLysIleAsnThr 140
QY 421 CAAAGAAGAACTCAAGTGTGAACCACTTCTGCAAACTCTGATACCAACAATAGGTAT 480
Db 141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160
QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAGAGGAGGTGTATACAGTTGAGGAAGCT 540
Db 161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluGluAla 180
QY 541 AAGAGTCTTGAACACTACATAGATGTGTTGAAGGAATGAAGTTGACCGTGTCTACCTC 600
Db 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyThrLeu 200
QY 601 TCTCCATCTTGTACTAATCTCTGAGAAAGTGTGTGCAACTGTGATACCTTATATC 660
Db 201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAlaIleThr 220
QY 661 CTTTGTGAATCAGAAAGATTTACTAGCATCAAGACATGCTACCAATCTTAGAACAGTT 720
Db 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAGTAAACCGTCCATCTCTTATTATGCTGAAGAGCTAGAAGGTGAAGCACTTCGT 840
Db 241 ValGlnThrGlyLysProLeuValIleAlaGluAspValGluGlyGluArgLeuAla 260
QY 781 ACACCTGTAGTCAATAGCTCGTGGAGCACTCCAAAGTGTAGCCGTAAAGCTCTCGT 840
Db 261 ThrLeuValValAsnLysLeuArgGlyLeuLysIleAlaAlaValLysAlaProGly 280
QY 841 TTTGTGAACCGCTAAAGCTATGCTTGAAGATATTGCTATCTCTACTGAGGAGAAGCA 900
Db 281 PheGlyAspArgGlyAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300
QY 901 ATATTGAGATCGTGGTATAAGCTTGAAGTGAAGTGAAGCTTGTCTTTTAGGAACAGCT 960
Db 301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyLysSer 320
QY 961 AAACGTGTAGTTATTGACAAAGAAATATCTATCTGTTGATGCTGCTGCAAAATCAGAA 1020
Db 321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCAGAAACAAAGCTCAGATTAT 1080
Db 341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluThrThrSerAspTyr 360
QY 1081 GATCGTGAAGAACTTCAAGAGCTCTGTGCAAACTTGTGTGGTGGAGTAGCTGTATCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380
QY 1141 GTTGAGCTGCTACTCAAACTGAAATGAAGAGAGAGATCGTGTAGAGATGCTCTA 1200
Db 381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgIleAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTGTGAAGAGGTATTGTCCCTGGTGGTGTACTGCTTTTTC 1260
Db 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420
QY 1261 CGTCCATTAAGTCTTCATGATATTAACCTGCTGATGATGAATCTGCTGACTT 1320

Db 421 ArgSerSerThrLysIle---ThrValLysGlyValAsnAspGlnGluAlaGlyIle 439
QY 1321 AATATCATCGTCTCTTCTTGAAGAGCCCTTTTACGTCAAAATTCCTCAATGCTGCTAT 1380
Db 440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459
QY 1381 GAAGGTTCCTATCTCTAGAAAAAGTTCTGTAACCA---AAAGATGCTTTTGATTAAAT 1437
Db 460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479
QY 1438 GCTGCATCAGGAAATATGAAGCCTTATTAAAGCTGGTGTCAATTCATCTCAAAAAGTT 1497
Db 480 AlaGlnThrGlyGlyLysLeuIleAlaLeuGlyIleValAspProValLysVal 499
QY 1498 ACAGGTATTGCATTACAAATCAGCATCAGTAGCTCTTACTTCTTACTACAGATGC 1557
Db 500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerValLeuIleThrThrGluAla 519
QY 1558 GCTATTGCTGAAAAACCAACCACTAAAAAGATATGCT---ATGCTGGCGGTGGTATG 1614
Db 520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyGlyMet 539
QY 1615 GGTGGTATG 1623
Db 540 GlyGlyMet 542

RESULT 14
B83098
GroEL protein P4385 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83098
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
gen.
A:Reference number: AB02950; MUID:20437337; PMID:10984043
A:Residues: 1-547 <STO>
A:Cross-references: GB:AE004854; GB:AE004091; NID:g9950606; PIDN:AAG07773.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: groEL; P4385
C:Superfamily: chaperonin groEL

Alignment Scores:
Pred. No.: 1,43e-92
Score: 1825.50
Percent Similarity: 80.51%
Best Local Similarity: 66.85%
Query Match: 61.51%
DB: 2

US-09-077-574A-1 (1-1647) x B83098 (1-547)

QY 1 ATGGCTCTTAAAGAAATCTTTTGTATGCTAAAGCCGTGAAACAACTTTCACAGGTGA 60
Db 1 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 20
QY 61 GATAAATCTGCAAACTCTTAAAGTAACACATTCAGCTTAAAGCCGTAAAGCCGTATTCGTTATT 120
Db 21 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGTCTTTTGGTTCCTCCAGTATTACAAAGATGGTGTATCTCTTCGCAAAAGAAAT 180
Db 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACCTGAAGATAGTTTGAATAATATGGCGCTCAAACTGTTAAAGAGTAGTCCCAAA 240
Db 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 80

```
QY 241 ACTAGCATATTGCTGGTGTAGTGAACCTACACAGCAACAGCTCCCTGACAGCTATTAT 300
Db      ::::::::::::::::::::
QY 81 AlaAsnAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaVal 100
Db      ::::::::::::::::::::
QY 301 CGTGAAGCTGTAAACCTGTAGCAGCTGGTGTAACTCTATGGCCATTAAACGTCGCATA 360
Db      ::::::::::::::::::::
QY 101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLe 120
Db      ::::::::::::::::::::
QY 361 GATAAAGCTGTTGTTGCTGTACTTAAAGAACTAAGACGACATTAACAAGCTTACTCGTAC 420
Db      ::::::::::::::::::::
QY 121 AspLysAlaThrValAlaAlaValAlaGlnLeuLysGluLeuAlaLysProCysAlaAsp 140
Db      ::::::::::::::::::::
QY 421 CAAAAAGAAATAGCTCAAGTGGTGAACCATTTCTGCAACCTGATACAACTAGTAAAT 480
Db      ::::::::::::::::::::
QY 141 ThrLysAlaAlaAlaGlnValGlyThrThrLeuAlaAsnSerAspGluSerIleGlyGln 160
Db      ::::::::::::::::::::
QY 481 ATCATAGCTGAAGCTATGGCTTAAAGTTGGAAGAGGAGGTGTATCACAGTTGAGGAAGCT 540
Db      ::::::::::::::::::::
QY 161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 180
Db      ::::::::::::::::::::
QY 541 AAAGCTCTTGAACCTACATTAGATGTGGTTGAAGAAATGAAGTTGACCGTGGCTACTCTC 600
Db      ::::::::::::::::::::
QY 181 SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
Db      ::::::::::::::::::::
QY 601 TCTCCATCTTTGTAACCTTAATCTCTGAGAAATGCTTTGTGAACCTGTATACCCCTATATC 660
Db      ::::::::::::::::::::
QY 201 SerProTyrPheValAsnLysProAspThrMetAlaAlaGluLeuAspSerProLeuLeu 220
Db      ::::::::::::::::::::
QY 661 CTTTGTATGAGAAAAAGATTACTAGCATGAAGACATGCTTACCAATCTTAGAACAAAGTT 720
Db      ::::::::::::::::::::
QY 221 LeuLeuValAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240
Db      ::::::::::::::::::::
QY 721 GCTAAAGTAAACCTCCACTCTTATTATCTGCTGAAGACGTAGAGGTGAAGCACTTGCA 780
Db      ::::::::::::::::::::
QY 241 AlalysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260
Db      ::::::::::::::::::::
QY 781 ACATTTGTAGTCATAGCTCCGTGGAGCACTCCAGTTGTAGCGGTAAAGCTCCCTGGT 840
Db      ::::::::::::::::::::
QY 261 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280
Db      ::::::::::::::::::::
QY 841 TTTGCTGAACCCGCTAAAGCTATGCTGAAGATATGCTATCTTACTGGAGGAGAACCA 900
Db      ::::::::::::::::::::
QY 281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 300
Db      ::::::::::::::::::::
QY 901 ATATTTGAAGATCGGTGATTAAGCTTAAAGCTTGAAGCTTGTCTTCTTTAGGACAGCT 960
Db      ::::::::::::::::::::
QY 301 IleSerGluGluValGlyLeuSerLeuGluGlyAlaThrLeuGluHisLeuGlyAsnAla 320
Db      ::::::::::::::::::::
QY 961 AAAGCTGTAGTATTGACAAAGAAATACTACTATCTGTTGATGCTGGTGAAGATCAGAA 1020
Db      ::::::::::::::::::::
QY 321 LysArgValValIleAsnLysGluAsnThrThrIleIleAspGlyAlaGlyValGlnAla 340
Db      ::::::::::::::::::::
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTAAGAAACAAAGCTCAGATTAT 1080
Db      ::::::::::::::::::::
QY 341 AspIleGluAlaArgValLeuGlnIleArgLysGlnIleGluGluThrThrSerAspTyr 360
Db      ::::::::::::::::::::
QY 1081 GATCGTGAAGAACTTCAAGAACGCTTTCGCAAACTTGTGTGGAGTAGTACTGTTATCCAT 1140
Db      ::::::::::::::::::::
QY 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleLys 380
Db      ::::::::::::::::::::
QY 1141 GTTGGAGCTGCTACTGAACCTGAATGAAGAAAGAGAGGATCGTGTAGAAGATGCTCTA 1200
Db      ::::::::::::::::::::
QY 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
Db      ::::::::::::::::::::
QY 1201 AATGCAACAAGAGCTGCGGTGAAGAAGGTATTGTCCTGGTGGTGGTACTGCTTTTCTC 1260
Db      ::::::::::::::::::::
QY 401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 420
Db      ::::::::::::::::::::
QY 1261 CGCTCCATTAAAGCTGTGATGATATTAAACCTGCTGATGATGAACCTGCTGGACTT 1320
Db      ::::::::::::::::::::
QY 421 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluGluGlnAsnValGlyIle 440
Db      ::::::::::::::::::::
```

```
QY 1321 AATATCATCGCTCGTCTCTTGAAGACGCTTTAGCTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db      ::::::::::::::::::::
QY 441 AlaLeuLeuArgArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyAsp 460
Db      ::::::::::::::::::::
QY 1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCTGGAACCAAAAGATGTTTGGATTAAATGCT 1440
Db      ::::::::::::::::::::
QY 461 GluProSerValValValAspLysValLysGlnGlySerGlyAsnTyrGlyPheAsnAla 480
Db      ::::::::::::::::::::
QY 1441 GCATCAGAGAAATATGAAGACCTTATTAAAGCTGCTGCTCATTTGATTCCTAAAAAGTTACA 1500
Db      ::::::::::::::::::::
QY 481 AlaThrGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 500
Db      ::::::::::::::::::::
QY 1501 CGTATTTCATTACAAAAATGCAGCATCAGTAGCTCCCTTACTTCTAACTCAGAAATGCGCT 1560
Db      ::::::::::::::::::::
QY 501 ArgSerAlaLeuGlnAlaAlaSerIleGlyLeuMetIleThrThrGluAlaMet 520
Db      ::::::::::::::::::::
QY 1561 ATTGCTGAAAAACCCAGAACCTTAAAGAAATATGCTATGCTGCGCGT-----GGTATG 1614
Db      ::::::::::::::::::::
QY 521 ValAlaGluIleValGlu-----AspLysProAlaMetGlyGlyMetProAspMet 537
Db      ::::::::::::::::::::
QY 1615 GCTGCTATGGTGGTATGAGCGGTATG 1641
Db      ::::::::::::::::::::
QY 538 GlyGlyMetGlyGlyMetGlyMet 546
Db      ::::::::::::::::::::
RESULT 15
S22347
groEL - Brucella abortus
C:Species: Brucella abortus
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 26-Aug-1999
C:Accession: S22347
R:Gor, D.; Mayfield, J.E.
B:Ochim. Biophys. Acta 1130, 120-122, 1992
A:Title: Cloning and nucleotide sequence of the Brucella abortus groE operon.
A:Reference number: S22346; MUID:92182006; PMID:11347461
A:Accession: S22347
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-546 <GOR>
A:Cross-references: EMBL:M82975; NID:gl44109; PIDN:AAA22997.1; PID:gl44111
C:Superfamily: chaperonin groEL
Alignment Scores:
Pred. No.: 1,96e-92 Length: 546
Score: 1823.00 Matches: 378
Percent Similarity: 80.51% Conservative: 64
Best Local Similarity: 68.85% Mismatches: 99
Query Match: 61.42% Indels: 8
DB: 2 Gaps: 5
US-09-077-574A-1 (1-1647) x S22347 (1-546)
QY 1 ATGCTCTTAAAGAAATCTTTTGTATGCTTAAAGCCGCTGAAACCTTTCACGAGGTGTA 60
Db      ::::::::::::::::::::
QY 1 MetAlaAlalysAspValLysPheGlyArgThrAlaArgGluLysMetLeuArgGlyVal 20
Db      ::::::::::::::::::::
QY 61 GATAAATCTTCAATGCTGTATAAGTAACTCGACCTTAAGCCGCTTAAGTCTGTTAT 120
Db      ::::::::::::::::::::
QY 21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValIle 40
Db      ::::::::::::::::::::
QY 121 GAAAGTCTTTTGTGTCCTCCAGTATTACAAAGATGCTGTATCTGTTGCAAAAGAAAT 180
Db      ::::::::::::::::::::
QY 41 GluLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluVal 60
Db      ::::::::::::::::::::
QY 181 GAATCTGAAGATAAGTTTGAATAATATGGCCCTCAATGGTTAAAGATGACTGCCCAAA 240
Db      ::::::::::::::::::::
QY 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80
Db      ::::::::::::::::::::
QY 241 ACTAGCATATTGCTGGTGTAGTGAACCTACACAGCAACAGCTCCCTGCAAGCTATTAT 300
Db      ::::::::::::::::::::
QY 81 ThrAsnAspThrAlaGlyAspGlyThrThrAlaThrValLeuGlyGlnAlaVal 100
Db      ::::::::::::::::::::
QY 301 CGTGAAGGTGTAAACCTTGTAGCAGCTGGTGTATCTATGCGCATTAACCGTGGCATA 360
Db      ::::::::::::::::::::
```

```
Db 101 GlnGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTCTTCTGCTGTTACTAAAGAACTAAGCGACATTACAAGCCTACTCGTGAC 420
Db 121 AspLeuAlaValAsnGluValValAlaGluLeuLeuLysAlaLysIleAsnThr 140
QY 421 CAAAAGAATAAGCTCAAGTTGAACCATTTCTGCAAACTCTGATACAAACAATAGGTAAT 480
Db 141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnAlaGluAlaGluIleGlyLys 160
QY 481 ATCATAGCTGAAGCTATGCTTAAGTTGAAAAGGAGGTGTTATCATCAGTTGAGAGCT 540
Db 161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180
QY 541 AAAGGTCTTGAACACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGCCTACCTC 600
Db 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyIleLeu 200
QY 601 TCTCCATCTTTGTAACTAATCTCTGAGAAAATGGTTGTGAACCTTGATACCCCTTATATC 660
Db 201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220
QY 661 CTTTGTAAAGTGAAGAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTT 720
Db 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaLeuLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCCCTATTATTGCTGAAGACGTAGAAGTGAAGCACTTGCA 780
Db 241 ValGlnThrSerLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTGTAGTCAATAAGCTCCGTGAGCAGCTCCAAGTTGTAGCCGTAAAGCTCCTGGT 840
Db 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGCTAAAGCTATGCTTGAAGATTGCTATCCTTACTGAGGAGAGCA 900
Db 281 PheGlyAspArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
QY 901 ATATTTGAAGATCGTGGTATAAAGCTTGAAGTTGAAGTTGTCTTCTTTAGCAACAGCT 960
Db 301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320
QY 961 AAACGTGTATTTATTCACAAAAGAAATACTACTATCTGTTGATGCTGCTGCAAAATCAGAA 1020
Db 321 LysLysValSerIleSerLysGluAsnThrIleValAspGlyAlaGlyGlnLysAla 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATCGTGCACAAATGCAAGAACCAAGCTCAGATTAT 1080
Db 341 GluIleAspAlaArgValGlyGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTGAACAACTTCAAGAACCTCTTGCACAACTTGTGGTGAAGTGTATCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleArg 380
QY 1141 GTTGGAGCTGCTACTCAAACTGAAATCAAGAGAAGAGATCGTGTAGAAGATGCTCTA 1200
Db 381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCTGCTGGTGGTACTGCTTTTTC 1260
Db 401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyThrAlaLeuLeu 420
QY 1261 CGC---TCCATTAAAGTCTTGATGATATTAAACCTGCTGATGATGAACCTTGCTGGA 1317
Db 421 ArgAlaSerThrLysIle-----ThrAlaLysGlyValAsnAlaAspGlnGluAlaGly 438
QY 1318 CTTAATATCATCGCTCTTCTTGAAGAGCCTTTTACGTCAAAATTCGTCAAATGCTGGC 1377
Db 439 IleAsnIleValArgAlaIleGlnAlaProAlaArgGlnIleThrThrAsnAlaGly 458
QY 1378 TATGAGGTCTTATGTTGTAGAAAAGTTCGTGAA---CCAAAAGATGTTTGCATT 1434
Db 459 GluGluAlaSerValIleValGlyLysIleLeuGluAsnThrSerGluThrPheGlyTyr 478
```

Search completed: January 28, 2004, 13:30:02
Job time : 57.5 secs

```
QY 1435 AATGCTGCATCAGCAGAAATATGAAGACCTTTATTAAAGCTGCTGCTATTGATCCTAAAAA 1494
Db 479 AsnThrAlaAsnGlyGluTyrGlyAspLeuIleSerLeuGlyIleValAspProValLys 498
QY 1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCCTCTTACTTCTTAACCTACAGAA 1554
Db 499 ValValArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuIleThrThrGlu 518
QY 1555 TGGCGTATTGCTGAAAAAACCCAGAACCTTAAAAAAGATATGCCTATGCCTGCGC-----GGT 1608
Db 519 AlaMetIleAlaGluLeu-----ProLysLysAspAlaAlaProAlaGlyMetProGly 536
QY 1609 GGTATCGGTGGTATGGTGGTATGGAC 1635
Db 537 GlyMetGlyGlyMetGlyMetAsp 545
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:28:52 ; Search time 60.5 Seconds
(without alignments)
11256.041 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 2968
Sequence: 1 agggcttaagaaatctct.....gtatggacggtatgtagtag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 1554272

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO.spool.p/US09077574/runat 28012004 125848 11734/app query.fasta_1.1799
-DB=published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09077574 @CGN 1 1 13 @runat 28012004 125848 11734
-NCPU=6 -ICPU=3 -NO.MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	1862.5	62.8	547	12	US-10-369-493-17524	Sequence 17524, A
2	1857	62.6	547	12	US-10-369-493-10014	Sequence 10014, A
3	1840.5	62.0	544	12	US-10-369-493-11760	Sequence 11760, A
4	1839.5	62.0	547	12	US-10-369-493-7917	Sequence 7917, Ap
5	1835.5	61.8	542	12	US-10-369-493-10295	Sequence 10295, A
6	1834	61.8	541	12	US-10-369-493-9287	Sequence 9287, Ap
7	1833	61.8	544	12	US-10-369-493-15468	Sequence 15468, A
8	1833	61.8	544	12	US-10-369-493-16217	Sequence 16217, A
9	1828.5	61.6	543	12	US-10-369-493-15265	Sequence 15265, A
10	1826	61.6	542	12	US-10-369-493-11922	Sequence 11922, A
11	1812.5	61.1	551	12	US-10-369-493-12330	Sequence 12330, A
12	1811	61.0	540	12	US-10-369-493-15838	Sequence 15838, A
13	1810	61.0	552	12	US-10-369-493-12170	Sequence 12170, A
14	1786	60.2	545	12	US-10-369-493-167	Sequence 167, App
15	1780.5	60.0	544	12	US-10-369-493-17132	Sequence 17132, A
16	1778.5	59.9	545	12	US-10-228-167A-2	Sequence 2, Appli
17	1771.5	59.7	543	12	US-10-369-493-11916	Sequence 11916, A
18	1770	59.6	536	12	US-10-369-493-4511	Sequence 4511, Ap
19	1770	59.6	540	12	US-10-369-493-7269	Sequence 7269, Ap
20	1746	58.8	545	15	US-10-192-419-2	Sequence 2, Appli
21	1746	58.8	548	11	US-09-415-849-1	Sequence 1, Appli
22	1746	58.8	548	11	US-09-276-455-10	Sequence 10, Appli
23	1743.5	58.7	544	12	US-10-369-493-23094	Sequence 23094, A
24	1737.5	58.5	547	12	US-10-369-493-10862	Sequence 10862, A
25	1721	58.0	521	12	US-10-369-493-9597	Sequence 9597, Ap
26	1720.5	58.0	546	12	US-09-882-227-418	Sequence 418, App
27	1716.5	57.8	540	12	US-10-369-493-181	Sequence 181, App
28	1716.5	57.8	547	15	US-10-046-649-2	Sequence 2, Appli
29	1710	57.6	544	9	US-09-841-132-400	Sequence 400, App
30	1708.5	57.6	548	12	US-10-369-493-513	Sequence 513, App
31	1708	57.5	544	12	US-10-289-762-153	Sequence 153, App
32	1698.5	57.2	542	12	US-10-369-493-17907	Sequence 17907, A
33	1693	57.0	551	12	US-10-369-493-20184	Sequence 20184, A
34	1683.5	56.7	641	15	US-10-267-311-51	Sequence 51, Appl
35	1658	55.9	551	12	US-10-369-493-18994	Sequence 18994, A
36	1650	55.6	544	10	US-09-759-272B-5	Sequence 5, Appli
37	1638.5	55.2	576	12	US-10-369-493-8322	Sequence 8322, Ap
38	1636	55.1	541	15	US-10-156-761-11465	Sequence 11465, A
39	1635	55.1	538	12	US-10-369-493-2950	Sequence 2950, Ap
40	1628.5	54.9	542	15	US-10-156-761-12526	Sequence 12526, A
41	1610.5	54.3	544	12	US-10-369-493-20276	Sequence 20276, A
42	1597	53.8	541	11	US-09-880-505-160	Sequence 160, App
43	1597	53.8	541	12	US-10-205-979-40	Sequence 40, Appl
44	1597	53.8	541	14	US-10-051-643-160	Sequence 160, App
45	1595.5	53.8	540	10	US-09-847-637B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-369-493-17524
; Sequence 17524, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17524
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-17524

Alignment Scores:

Pred. No.: 6,45e-144 Length: 547
 Score: 1862.50 Matches: 366
 Percent Similarity: 81.39% Conservative: 80
 Best Local Similarity: 66.79% Mismatches: 97
 Query Match: 62.75% Indels: 5
 DB: 12 Gaps: 2

US-09-077-574A-1 (1-1647) x US-10-369-493-17524 (1-547)

QY 1 ATGGCTTCTAAGAAATCCTTTTGTGCTAAAGCCCGTGAAGAAATCTTACGAGGTGTA 60
 Db 1 MetAlaAlaLysGluLeuIlePheSerGluLysAlaArgSerMetValHisGlyVal 20
 QY 61 GATAAACTTGAACCTGCTTAAAGTAAACACTTGGACCTAAAGCCGTATGCTGTTAT 120
 Db 21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgHisValValLeu 40
 QY 121 GAAAGTCTTTTGGTTCCTCCAGTATTACAAAAGATGGTGTATCTGTGTCAAAAGAAAT 180
 Db 41 AspLysSerPheGlySerProIleThrLysAspGlyValSerValAlaLysGluLeu 60
 QY 181 GAACTTGAACATAGTTGAAATATATGGCGCTCAAAATGTTTAAAGTAGCTCCAAA 240
 Db 61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80
 QY 241 ACTPAGCATATTGCTGTGTATGAACTTACAAACAGCAACAGCTCTTGCACAAAGCTATTAT 300
 Db 81 ThrAsnAspHisAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaLeuLeu 100
 QY 301 CGTGAAGGTGTAACCTTGPAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 Db 101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLeu 120
 QY 361 GATAAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 Db 121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysLysLysLysLysLysLysLys 140
 QY 421 CAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATCAACAAATAGTAAAT 480
 Db 141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160
 QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGGAAGAGGAGGTGTTATCAGTTGAGGAAGCT 540
 Db 161 IleAlaAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGly 180
 QY 541 AAAGTCTTGAACACTACATAGTATGCTTGAAGGAATGAAGTTTGAAGCTGCTGCTACCT 600
 Db 181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200
 QY 601 TCTCCATCTTTGTAACCTAATCTGAGAAATCGTTTGTGAACCTTGTATCAACCTTATATC 660
 Db 201 SerProTyrPheIleAsnAsnGlnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220
 QY 661 CTTTGTAAAGTGAAGAAAGATTACTAGCATGAAAGACATCTACCAATCTTGAAGAAAGTT 720
 Db 221 LeuLeuHisAspLysLysIleSerSerValArgAspLeuLeuThrValLeuAspAlaVal 240
 QY 721 GCTAAAGTAAACCGTCCCTCTTATTGCTGGAACAGCTAGAGCTGAGAAGCTGACCTGCA 780
 Db 241 AlalysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260
 QY 781 ACACCTTGATGCTCAATAGCTCCGTGAGACCTCCCAAGTTGTAGCCGTAAAGCTCTGCT 840
 Db 261 ThrLeuValValAsnAsnIleArgGlyIleLysValCysAlaValLysAlaProGly 280
 QY 841 TTTGTTGAACCGCTGAAGCTATGCTTGAAGTATTGCTTATCTTACTCTGAGGAGCA 900
 Db 281 PheGlyAspArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300
 QY 901 ATATTGAGATCGTGTATAGCTTGAAGTGAAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960
 Db 301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrThrSerHisLeuGlyLysAla 320

QY 961 AAACGTGCTAGTATTGACAAAGAAATCTACTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 321 LysLysValArgValSerLysGluAsnThrThrIleLeuAspGlyIleGlyAspAsnAsp 340
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAGAAACAAAGCTCAGATTAT 1080
 Db 341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360
 QY 1081 GATCGTGAAGAACTTCAAGAACTCTTGTGAAAACCTTGTGCTGAGTAGCTGTTATCCAT 1140
 Db 361 AspArgGluLysLeuGlnArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
 QY 1141 GTTGGAGCTGCTACTGAACTGAAATCAAGAGAAAGAGATCGCTGAGAGATGCTCTA 1200
 Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAlaLeu 400
 QY 1201 AATGCAACAGAGCTCGGTTGAAGAGAGTATTGCTCCCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420
 QY 1261 CGCTCCATTAAAGTCTTGTATGATATTAAACCTGCTGATGATGATGATGATGATGATGAT 1320
 Db 421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyValaAsnGluAspGlnThrHisGlyIle 440
 QY 1321 AATATCATCGTCTGCTTCTTGAAGAGCCCTTTACGTCAAATTCGCTCAAAATGCTGCTAT 1380
 Db 441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValaAsnAlaGlyGlu 460
 QY 1381 GAAGGTCTTATTGTTAGAAAAGTTTCGTGAACCAAAAGATGCTTTGATTTAATGCT 1440
 Db 461 GluProSerValIleLeuAsnLysValLysGluLysAspAsnPheGlyTyrAsnAla 480
 QY 1441 GCATCAGGAGATATGAACCTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 481 AlaThrGlyGluPheGlyAspMetValAsnLeuGlyIleLeuAspProThrLysValThr 500
 QY 1501 CGTATTGCTTCAAAATCGACATCAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 Db 501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520
 QY 1561 ATTGCTGAAAAACCAAGCACTTAAAGATATGCTTATGCTGCTGCTGCTGCTGCTGCTGCT 1611
 Db 521 ValAlaGlu-----AlaProLysLysAspGluProThrProAlaAlaGlyGlyGly 538
 QY 1612 ATCGGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1635
 Db 539 MetGlyGlyMetGlyGlyMetAsp 546

RESULT 2

US-10-369-493-10014
 ; Sequence 10014, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 10014
 ; LENGTH: 547
 ; TYPE: PR1
 ; ORGANISM: magnetite-containing magnetic coccus
 US-10-369-493-10014

Alignment Scores:

Pred. No.: 1,82e-143 Length: 547
Score: 1857.00 Matches: 382
Percent Similarity: 82.55% Conservative: 72
Best Local Similarity: 69.45% Mismatches: 90
Query Match: 62.57% Indels: 6
DB: 12 Gaps: 4

US-09-077-574a-1 (1-1647) x US-10-369-493-10014 (1-547)

QY 1 ATGGCTTCTAAAGAAATCTCTTTTGTAGCTAAAGCCCGTGAAGAACTCTTCAAGAGTGTA 60
Db 1 MetAlaAlaLysGluValLysPheGlyGluAlaAlaArgAlaLysMetLeuAsnGlyVal 20
QY 61 GATAAATCTGCAATGCTGTTAAAGTAACTTGGACCTTAAGCGCGTATGCTGTTATT 120
Db 21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValLeu 40
QY 121 GAAAGCTTTTGGTCCCGAGTATTACAAAGATGGTGTATCTGTGTCAGAAAGAAATT 180
Db 41 AspLysSerTrpGlyAlaProArgMetThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAATCTGAAGATAAGTTTGAATATATGGCGCTCAAAATGGTTAAAGAGTAGCTCCAAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValSerLys 80
QY 241 ACTAGCGATATCTGCTGATGGAATCTAACAGCAACAGTCTCTTGCACAAAGCTATTAT 300
Db 81 ThrAlaAspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIle 100
QY 301 CGTGAAGGTCTAAACTTGTAGCAGCTGGTGGTAACTCTTATGCGCTTAAAGCTGGCATA 360
Db 101 ArgGluGlyMetLysAlaValAlaAlaGlyMetAsnProMetAspLysValArgGlyIle 120
QY 361 GATAAAGCTCTTGTGCTGTACTTAAAGAACTAAGCGACATTAAGACCTACTCTCGTAC 420
Db 121 AspLeuAlaValGluAlaValValValGlyLeuLysGlyIleSerArgGluValAlaAsn 140
QY 421 CAAAGAAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGATACCAAGTAGGTAAT 480
Db 141 SerGlnGluIleAlaGlnValGlyAlaIleSerAlaAsnSerAspLysValValGlyAsp 160
QY 481 ATCATAGCTGAAGTATGCTTAAAGTAAAGAGGAGTGTATCAAGTTGGAGGAGCT 540
Db 161 MetIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluGluAla 180
QY 541 AAAGTCTTCAAACTACATTAGATGCTGTGAAGAAATGAAGTTGACCTGGCTACCTC 600
Db 181 LysGlyLeuGluThrThrLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATACCTTGTAACTAATCTCTGAGAAATGGTTTGTGAACCTGTATATACCTTATATC 660
Db 201 SerProTyrPheValThrAsnAlaAspLysMetLeuValGlnMetGluAsnProLeuIle 220
QY 661 CTITGTAAATGAGAAAGATTAAGTATGCAAGACATGCTTACCACTTAGAACAAAGTT 720
Db 221 LeuLeuValGluLysLysIleSerAsnLeuGlnIleLeuGlnIleLeuGluGlyAla 240
QY 721 GCTAAAGTAAACCTCCACTCTTATTATTGCTGAAGACGTAGAGGTGAAGCACTTGCA 780
Db 241 ValGlnSerSerArgProLeuMetIleIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTTAGTCAATAAGCTCCGTGGAGCACTCCAAAGTTGAGCGGTAAAGCTCTCGGT 840
Db 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysValCysAlaValLysAlaProGly 280
QY 841 TTTGCTGAACCGCTTAAGCTATCTTGAAGATATGCTATCTTACTGAGGAGGAGCA 900
Db 281 PheGlyAspArgLysAlaMetMetGluAspIleAlaThrLeuThrGlyGlyValLeu 300
QY 901 ATATTTGAAGATCGTGTATAAAGCTGAAATGTAAAGCTTGTCTTCTTTAGGAACAGCT 960
Db 301 ValSerGluAspValGlyValLysLeuGluAsnValThrMetAspMetLeuGlyMetAla 320

QY 961 AAACGTGTAGTTATTGACAAAGAAATATCTATCTATCTGTTGATGCTGCTGAAATCAGAA 1020
Db 321 LysSerIleValValThrLysGluAspThrThrIleIleAspGlyAlaGlyAspHisGlu 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGACAAATTCGAGAAACAACTCAGATTAT 1080
Db 341 AlaIleLysAlaArgValAsnGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360
QY 1081 GATCGTGAAGAACTTCAAGAACTCTTGTCAAACTTGTGCTGAGTAGTGTATTCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaIleLys 380
QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGCAAGAGGATCGCTAGAGATGCTCTA 1200
Db 381 ValGlyGlyAlaThrGluValGluValLysGluArgLysAspArgValAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGTATTGTCCCTGGTGGTACTGCTTTTGTGTC 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyValAlaLeuLeu 420
QY 1261 CGCTCCATTAAAGTCTTCTTCTTGAAGAGCTTTTACGTCAAAATTCGTCAAAATGCTGGC 1317
Db 421 ArgAlaArgGluAlaSerLeuThrAsnLeuGlnGlyAlaAsnHisAspGlnGlnValGly 440
QY 1318 CTTAATATCATCGTCTGCTTCTTGAAGAGCTTTTACGTCAAAATTCGTCAAAATGCTGGC 1377
Db 441 IleAsnIleValThrArgAlaLeuGluGluProLeuArgIleIleAlaSerAsnAlaGly 460
QY 1378 TATGAAGGTTCTATTGTTGTAAGAAAGTTCGTGAACCAAAAGAT---GGTTTGGATT 1434
Db 461 AlaGluGlySerValValValAsnArgValValGluThrLysGluThrAsnPheGlyPhe 480
QY 1435 AATGCTGCATCAGAGATATGAAGCTTATTAAAGCTGCTGCTCATGATGCTTAAATAA 1494
Db 481 AsnAlaAlaThrGlyValTyrGluAspLeuValAlaSerGlyValIleAspProAlaLys 500
QY 1495 GTTACAGTATTTCATTTACAAATGACAGCATGAGTACCTCTTCTTCTTAACTTACAGAA 1554
Db 501 ValValArgHisAlaLeuGlnAlaAlaAlaSerValAlaGlyLeuMetIleThrThrGlu 520
QY 1555 TGGCTATTCTGCTCAAAACCAAGAACCTTAAAGATATGCT---ATGCTGCGGTGGT 1611
Db 521 AlaMetValAlaGluLeuPro-----LysAspGluProAlaMetProGlyGlyAsp 537
QY 1612 ATGCGTGGTATGGTGGTATGACGGTATG 1641
Db 538 MetGlyGlyMetGlyGlyMetGlyMetGlyMet 547

RESULT 3

US-10-369-493-11760
; Sequence 11760, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11760
; LENGTH: 544
; TYPE: PRF
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11760

Alignment Scores:

Pred. No.: 4,09e-142 Length: 544
 Score: 1840.50 Matches: 376
 Percent Similarity: 81.95% Conservative: 69
 Best Local Similarity: 69.24% Mismatches: 95
 Query Match: 62.01% Indels: 3
 DB: 12 Gaps: 3

US-09-077-574A-1 (1-1647) x US-10-369-493-11760 (1-544)

```

QY 1 ATGGCTCTTAAGAAATCCTTTTGTATGCTAAAGCCCGTGAATAAATCTTACAGAGTGTA 60
Db 1 MetAlaAlaGluValLysPheGlyArgSerAlaArgGluLysMetLeuLysGlyVal 20
QY 61 GATAAACTTGCAATCTGTTAAAGTAACACTTGGACCTAAAGCCGCTAATGCGTTATT 120
Db 21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40
QY 121 GAAAAGTCTTTGGTTCCTCCAGTTATTACAAAAGATGGTGTATCTCTTGCACCAAGAAAT 180
Db 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACCTGAAGATAGTTGAAAATATGGCGGCTCAAAATGGTTAAAGAGTAGTCCCAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80
QY 241 ACTAGCATATTGCTGCTGATGGAATCAACAGCAACAGCTCTCTGCACAGCTATTAT 300
Db 81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CTGAAGGTGTAACCTTGTAGCAGTGGTCTGTAATCTTATGGCCATTAAAGTGGCATA 360
Db 101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAGCTTGTGCTGCTTACTAAAGAACTAAGCAATTAAGCAATTAAGCAATTAAG 420
Db 121 AspLeuAlaValAlaGluValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr 140
QY 421 CAAAAGAAATAGCTCAAGTTCGAACCATTTCTGCAAACTCTCATACACATAGTANT 480
Db 141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTGGAAAGAGAGTGTGTATCACAGTTGAGGAAGCT 540
Db 161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluAla 180
QY 541 AAAGTCTTGAACCTACATAGATGCTGTTGAAGNATGAAGTTGACCGTGGCTACCTC 600
Db 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATCTTTGTAACTAATCTGAGAAATGGTTGTGAACCTTGATACCTTATATC 660
Db 201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220
QY 661 CTTTGTAAAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAAACAAGTT 720
Db 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATTTGCTGAAGACGTAGAGGTGAAGCACTTGA 780
Db 241 ValGlnThrGlyLysProLeuValIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACACCTTGAGTCAATTAAGCTCCGTGGAGCACTCAAGTTGTAGCCGTAAAGCTCTGCT 840
Db 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280
QY 841 TTTGTTGAACCGCTGAAGCTATGCTTGAAGATATTGCTTATCTTACTTGGAGGAGAACA 900
Db 281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300
QY 901 ATATTGAGATCGGTGATAAGCTTGAAGATGAAGCTTCTCTCTTCTTCTAGCAAGCT 960
Db 301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetIleuGlyLysSer 320

```

```

QY 961 AAACGTGTAGTTATTGACAAAAGAAATACTACTATCTGTTGATGCTGCGAAATCAGAA 1020
Db 321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGACAAATTCAGAAAACAGCTCAGATTAT 1080
Db 341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTGAATAACTTCAAGAACGCTCTTGTGAAAACCTTGTGTTGGTGGAGTAGCTGTATCCAT 1140
Db 361 AspArgGluLysLeuGlnArgLeuAlaLysLeuAlaGlyValAlaValIleArg 380
QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAGAAAGCAAGCATCGCTGAGAGATGCTCTA 1200
Db 381 ValGlyGlySerThrGluValGluValLysGluLysAspArgIleAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCTCGTGGTGGTATCTGCTTTGTC 1260
Db 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420
QY 1261 CGCTCCATTAAAGTCTTCATGATATTAACCTGCTGATGATGATGATGATGATGATGAT 1320
Db 421 ArgSerSerThrLysIle---ThrValLysGlyValAsnAspGlnGluAlaGlyIle 439
QY 1321 AATATCATCGTCTGCTCTTGAAGAGCCCTTTACGTCAAATTCGCTCAATTCGCTGCTAT 1380
Db 440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459
QY 1381 GAAGGTTCTATTGTTAGAAAAGTTCTGTGAACCA---AAAGATGGTTTGGATTAAAT 1437
Db 460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479
QY 1438 GCTGATCAGGAGATATGAACCTTATAAGCTGTGTGCTGATGATGATGATGATGATGAT 1497
Db 480 AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499
QY 1498 ACACGTATTGCTATCAAAATCGACATCGACATCGACATCGACATCGACATCGACATCG 1557
Db 500 ValAlaGThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 519
QY 1558 GCTATTGCTGAAAACCAAGCAACCTAAAGAAAGATATGCTCT---ATGCTCGCGGTGATG 1614
Db 520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyMet 539
QY 1615 GTGCTATG 1623
Db 540 GlyGlyMet 542

```

RESULT 4

US-10-369-493-7917
 ; Sequence 7917, Application US/10369493
 ; Publication NO. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 7917
 ; LENGTH: 547
 ; TYPE: PRT
 ; ORGANISM: Rhodobacter sphaeroides
 US-10-369-493-7917

Alignment Scores:

Pred. No.: 4, 95e-142 Length: 547
 Score: 1839.50 Matches: 372
 Percent Similarity: 80.66% Conservative: 70
 Best Local Similarity: 67.88% Mismatches: 103
 Query Match: 61.98% Indels: 3
 DB: 12 Gaps: 2

US-09-077-574A-1 (1-1647) x US-10-369-493-7917 (1-547)

QY 1 ATGCTTCTAAAGAAATCTTTTGTAGCTAAAGCCGCTGAAAACATTTTACGAGGTGA 60
 Db 1 MetAlaAlaLysAspValLysPheAspThrAspAlaArgAspArgMetLeuArgGlyVal 20
 QY 61 GATAAAGCTTCAATGCTGTTAAAGTAACTTGGAGCTTAAAGCCCTAATGCTGTTATT 120
 Db 21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValle 40
 QY 121 GAAAAGCTTTTGGTCCCGAGTTATTACAAAAGATGCTGTATCTGTTGCAAAAAGAAAT 180
 Db 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60
 QY 181 GAACTTGAAGATAGTTTGAATATATGGCGCTCAAAATGGTTAAAGAGTAGCTCCAAA 240
 Db 61 GluLeuSerAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80
 QY 241 ACTAGCATATTGCTGTGATGAGTAACTACACAGCAACAGTCTCTTGCACAGCTATTAT 300
 Db 81 ThrAsnAspGluAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIle 100
 QY 301 CGTGAAGCTGTAAGCTTTAGACAGCTGCTGTAATCTTATGGCCATTAAACGTTGGCATA 360
 Db 101 LysGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLe 120
 QY 361 GATAAAGCTTGTGCTGTTACTTAAAGAACTAAGCAGCATTAACAAGCCCTACTCGTAC 420
 Db 121 AspLeuAlaThrSerLysValValGluAlaIleLysAlaAlaArgProValAsnAsp 140
 QY 421 CAAGAAGAAATAGCTCAAGTTGGACCATTTCTGCAAACTCTGATACACAACTAGTAA 480
 Db 141 SerHisGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluAlaGlnIleGlyArg 160
 QY 481 ATCATAGCTGAAGCTATGGCTTAAAGTTGGAAGAGGCTGTATCACAGTTTGAGGAAGCT 540
 Db 161 PheIleAlaAspAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAsn 180
 QY 541 AAAGCTCTTGAACATACATTAGATGTGTTGAAGAAATGAAGTTTGACCGTGGCTACCTC 600
 Db 181 LysGlyLeuGluThrGluValGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
 QY 601 TCTCCATCTTTGTAATCTCTGAGAAATGTTTGTGAACCTTGATACCCCTTATATC 660
 Db 201 SerProTyrPheValThrAsnAlaAspLysMetThrAlaGluLeuAspAspValTyrIle 220
 QY 661 CTTTGTATGAGAAAGATGACTAGCAATGAAAGACATGCTACCACTCTTAGAACAAGTT 720
 Db 221 LeuLeuHisGluLysLysLeuSerSerLeuGlnProMetValProLeuLeuGluAlaVal 240
 QY 721 GCTAAAGTAAACCGCTCACTCTTATTATTGCTGAAGACGTAGAGGTGAAGCACTTGCA 780
 Db 241 IleGlnSerGlnLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260
 QY 781 ACATTGTAGTCAATAAGCTCCGTGGAGCATCTCAAGTTGTAGCGGTAAAGCTTCCTGGT 840
 Db 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280
 QY 841 TTTGTGAAACCGCTTAAAGCTATGCTTGAAGATATGCTATCTTACTTGGAGGAGCA 900
 Db 281 PheGlyAspArgGlyLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyGlnVal 300
 QY 901 ATATTGAGATCGTGGTATAAAGCTTGAAGATGTAAGCTTGTCTTCTTTAGGACAGCT 960
 Db 301 IleSerGluAspLeuGlyMetLysLeuGluAsnValThrIleAspMetLeuGlyArgAla 320

QY 961 AAACGTGTAGTATTTCACAAAGAAATATCTACTATCTGTTGATGCTGCTGAAATACGAA 1020
 Db 321 LysLysIleSerIleAsnLysAspAsnThrThrIleValAspGlyAsnGlyAspLysAla 340
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCGCAAAATTCGAAGAAACAAGCTCAGATT 1080
 Db 341 GluIleAspAlaArgValAlaGlnIleArgAsnGlnIleGluGluThrSerSerAspTyr 360
 QY 1081 GATCGTGAAGAACTTCAGAACCTCTTGCAAAACCTTCTTGGTGAGTAGCTGTTATCCAT 1140
 Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleArg 380
 QY 1141 GTTGGAGCTCTACTGAACTGAAATGAAAGAGAAAGAGATCGTGTAGAGATGCTCTA 1200
 Db 381 ValGlyGlyMetThrGluValGluValLysGluArgLysAspArgValAspAspAlaLeu 400
 QY 1201 AATGCAACAAGAGCTCGGTTGAAGAGGTATTGTCTCCCTGGTGGTGTACTGCTTTTGTG 1260
 Db 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValValGlyGlyGlyValAlaLeuIle 420
 QY 1261 CGTCCATTAAAGCTTGTGATGATATTAAACCTGCTGATGATGATGATGATGATGATG 1320
 Db 421 GlnGlyGlyLysAlaLeuAspGlyLeuThrGlyGluAsnProAspGlnAsnAlaGlyIle 440
 QY 1321 AATATCATCGTCTGCTCTTGAAGAGCTTTTACGTCAAATTCGTCAAATTCGTGCTAT 1380
 Db 441 ThrIleValArgArgAlaLeuGluAlaProLeuArgGlnIleAlaGlnAsnAlaGlyVal 460
 QY 1381 GAAGGTTCTATTGTTGTAGAAAAGTTTCGTCAACCAAAAGAT---GGTTTGGATTAA 1437
 Db 461 AspGlySerValValAlaGlyLysValArgGluSerAsnGluLysSerPheGlyPheAsn 480
 QY 1438 GCTGCATCAGAGAAATATGAGACCTTATTAAAGCTGCTGTCATGATGATGATGATGAT 1497
 Db 481 AlaGlnThrGluGluTyrGlyAspMetPheLysPheGlyValIleAspProAlaLysVal 500
 QY 1498 ACAGTATTGCTATTACAAATGACAGCATCAGTACCTCTTACTTCTTAACCTACAGAATGC 1557
 Db 501 ValArgThrAlaLeuGluAspAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 520
 QY 1558 GCTATTGCTGAAAACCAAGACCTTAAAGAAATATGCTATGCTGCTGGCGGTGGTATGG 1617
 Db 521 MetIleAlaAspLysProGluProLysSer-----ProAlaGlyGlyProGlyMetGly 538
 QY 1618 GGTATGGTGTATGAGCGGTATG 1641
 Db 539 GlyMetGlyGlyMetAspGlyMet 546

RESULT 5

US-10-369-493-10295
 ; Sequence 10295, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 10295
 ; LENGTH: 542
 ; TYPE: PRT
 ; ORGANISM: Cytophaga hutchinsonii
 US-10-369-493-10295

Best Local Similarity: 66.30% Mismatches: 101
Query Match: 61.79% Indels: 2
DB: 12 Gaps: 1

US-09-077-574A-1 (1-1647) x US-10-369-493-9287 (1-541)

```
QY 1 ATGGCTTCTAAAGAAATCTTTTATGCTAAAGCCCGTGAAGAAATCTTACGAGGTGTA 60
DB 1 MetAlaAlaLysGluLeuIlePheSerGluLysAlaArgSerArgMetValTyrGlyVal 20
QY 61 GATAAAGCTGCAATGCTGTTTAAAGTAACTTGAACCTTGAACCTGAGCCGCGTAAATGCTGTTAT 120
DB 21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgHisValValLeu 40
QY 121 GAAAGCTTTTGGTCCCGAGTATTAACAAAGATGCTGATCTGTTGCAAGAAAT 180
DB 41 AspLysSerPheGlySerProIleThrLysAspGlyValSerValAlaLysGluLeu 60
QY 181 GAACCTTGAAGATAAGTTTGAAGATATGGCGCTCAATGGTAAAGAGTACCTCCCAAA 240
DB 61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTGCTGTGATGGAACCTACAAAGCAACAGCTCTTGCACAAAGCTATTAT 300
DB 81 ThrAsnAspHisAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuLeu 100
QY 301 CGTGAAGGTGTAACCTTGTGCTGTTACTAAAGACATTAAGCGACATTAAGAGCTTACTCGTGAC 420
DB 101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLeu 120
QY 361 GATAAAGCTGTTGTTGCTGTTTACTAAAGACATTAAGCGACATTAAGAGCTTACTCGTGAC 480
DB 121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysIleSerLysProThrSerAsp 140
QY 421 CAAAAGAAATPAGCTCAAGTTGGAACCATTTCTGCAACTCTGATACAAACATAGGTAAT 480
DB 141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160
QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAGAGAGTGTATACAGATTGAGGAAGCT 540
DB 161 IleIleAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGluGly 180
QY 541 AAAGGTCTTGAACCTACATTAAGTGTGTTGAAGAAATGAAGTTTGAACGCTGCTACCTC 600
DB 181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200
QY 601 TCTCCATACCTTTGTAACCTAATCTCTGAGAAATGTTTGTGAACCTTGATTAACCTTATATC 660
DB 201 SerProTyrPheIleAsnAsnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220
QY 661 CTTTGTATGAGAAAGAAATGATAGCATGATGAAGCATGCTACCAATCTTAGAACAAGTT 720
DB 221 LeuLeuHisAspLysLysIleSerSerValArgAspLeuLeuThrValLeuAspAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATGCTGAAGACGTAGAAGGTGAAGCACTTCCA 780
DB 241 AlalysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260
QY 781 ACACCTGTGAGTCAATAGCTCCGAGCACTCCAGTGTAGCGGTGAAGAAAGCTCTGGT 840
DB 261 ThrLeuValValAsnAsnIleArgGlyIleIleLysValCysAlaValAlaLysAlaProGly 280
QY 841 TTTGGTGAACCGGTAAAGCTATGCTTGAAGATATGCTATCTCTTACTTGGAGGAGCA 900
DB 281 PheGlyAspArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300
QY 901 ATATTGAGATCGTGTATTAAGCTTCAAAATGTAAGCTGTCTCTTTTAGGACACGT 960
DB 301 IleSerGluGluValGlyLeuSerLysGluLysAlaThrThrSerHisLeuGlyLysAla 320
QY 961 AAAGCTGTAGTATTGACAAAGAAATACTATATCGTTGATGCTGCTGGAATATCAGAA 1020
DB 321 LysLysValArgValSerLysGluAsnThrThrIleIleAspGlyIleGlyAspAsnAsp 340
```

```
QY 1021 GATATTAAAGCTCGAGTTAAACAATTCGTGCACAAATTTGAAGAAACAAGCTCAGATTAT 1080
DB 341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTGAAGAAATCTCAAGAACGCTTCGCAAAATCTTGTGTGGAGTAGCTGTTATCCAT 1140
DB 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTGGAGCTGCTACTGAAACTGAATGAAGAGAGAAAGAGATCGTGTAGAAGATGCTCTA 1200
DB 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAlaLeu 400
QY 1201 AATGCAACAAGAGCTGCGGTTGAAGAGGTATTCCTCGTGTGGTGGTACTGCTTTGTC 1260
DB 401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuLeu 420
QY 1261 CGTCCCATTTAAAGTCCTTGATGATATTAACCTGCTGATGATGATGAACCTTGTGGACTT 1320
DB 421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyAlaAsnGluAspGlnThrHisGlyIle 440
QY 1321 AATATCATCCGCTGCTTCTTGAAGAGCCTTTAGCTCAATGCTGCAAAATGCTGGCTAT 1380
DB 441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460
QY 1381 GAAGGTCTATTGTTGTAGAAAAGTTCGTGAACCAAGAGATGTTTGGATTTAATGCT 1440
DB 461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnPheGlyTyrAsnAla 480
QY 1441 GCATCAGAGAAATATGAAGACCTTATTAAAGCTGCTGATTCCTTAAAGAAAGTTTACA 1500
DB 481 AlaThrGlyGluPheGlyAspMetValLysLeuGlyIleLeuAspProThrLysValAla 500
QY 1501 CGTATTGCTATTACAAATGCAAGATCAGATCAGTCCCTTACTTCTTAACCTACAGATGCGCT 1560
DB 501 ArgThrAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520
QY 1561 ATTCTCTGAAGAAACCAAGAACCTTAAAGAAATATGCTATGCTGCGGCTGGTATGCGTGT 1620
DB 521 ValAlaGlu-----AlaProLysLysAspGluProThrProAlaAlaGlyGlyGly 538
QY 1621 ATGGGTGCT 1629
DB 539 MetGlyGly 541
RESULT 7
US-10-369-493-15468
; Sequence 15468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15468
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15468
Alignment Scores:
Pred. No.: 1,68e-141 Length: 544
Score: 1833.00 Matches: 365
Percent Similarity: 81.87% Conservative: 82
```

Best Local Similarity:	66.85%	Mismatches:	95
Query Match:	61.76%	Indels:	4
DB:	12	Gaps:	3

US-09-077-574A-1 (1-1647) x US-10-369-493-15468 (1-544)

Qy	1	ATGCTCTCTAAAGAAATCCCTTTTGTAGTGTAAAGCCGTGAAAAACCTTTACAGAGGTGA	60
Db	1	MetAlaAlaLysAspIleArgPheGlyGluAspAlaArgThrArgMetValArgGlyVal	20
Qy	61	GATAAACTTCGAATGCTGTTAAAGTAACACTTGGACCTAAAGCCGTAAATGCTGTTAT	120
Db	21	AsnValLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu	40
Qy	121	GAAGAAGTCTTTTGGTTCCTCCAGTTATTACAAAAGATGGTGTATCTGTGTCGAAAGAAAT	180
Db	41	GluLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle	60
Qy	181	GAACCTGAAGTAAGTTGAAAAATATGGCGCTCAATGGTTAAAGAGTAGCTCCCAA	240
Db	61	GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg	80
Qy	241	ACTAGCGATATTGCTGCTGTGATGGAACTACACACACAGCTCCCTGCACCAAGCTATTAT	300
Db	81	ThrAsnAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaLeuIle	100
Qy	301	CGTGAAGCTGTAAACCTTGTAGCAGCTGGTGGTAAATCCTATGGCCATTAAACGTGGCATA	360
Db	101	ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle	120
Qy	361	GATAAAGCTGTGTGCTGTACTTAAAGMACTAAGCACATTAACAAGCCCTACTCTGCAC	420
Db	121	AspGlnAlaValLysAlaAlaValIleGluLeuLysAsnIleSerLysProThrThrAsp	140
Qy	421	CAAAAAGAAATAGCTCAAGTTGGAAACCATTTCTCAAACTCTGATCAACAATAGGTAAT	480
Db	141	AspLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn	160
Qy	481	ATCATAGCTGAAGCTATGGCTPAAAGTTGGAAAGAGGAGTGTTATCATCAGTTTGAGGAAGCT	540
Db	161	IleIleAlaGluAlaMetGlnLysValGlyLysGluGlyValIleThrValGluGluGly	180
Qy	541	AAAGGCTCTTGAACTACATATTAGATGCTGGTGAAGGAATGAAGTTGACCGTGCTACCTC	600
Db	181	SerGlyLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyIle	200
Qy	601	TCTCCATACTTTGTAACCTAATCTCTGAGAAATGGTTGTGAACCTGATAACCCCTATATC	660
Db	201	SerProTyrPheIleAsnAsnGlnSerGlnSerAlaAspLeuAspAspPropheile	220
Qy	661	CTTTGTAATGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT	720
Db	221	LeuLeuHisAspLysLysIleSerAsnValArgAspLeuLeuProValLeuGluGlyVal	240
Qy	721	GCTAAAGTAAACCGTCCACTCCTTATTATTCTCAAGACGTAGAGGTGAAGCACTTGCA	780
Db	241	AlaLysAlaGlyLysProLeuLeuIleValAlaGluGluValGluGluAlaLeuAla	260
Qy	781	ACACTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTGGT	840
Db	261	ThrLeuValValAsnThrIleArgGlyIleValLysValValAlaValLysAlaProGly	280
Qy	841	TTTGGTGAACCGGTAAAGCTATGCTTGAAGATATTGCTATCTCTTACTGGAGGAGACCA	900
Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal	300
Qy	901	ATATTGGAAGTCTGGTGTATAAAAGCTTGAAAAATCTTAAGCTTGTCTTCTTAGAACAGCT	960
Db	301	IleSerGluGluValGlyLeuAlaLeuGluLysAlaThrIleLysAspLeuGlyArgAla	320
Qy	961	AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGTGCTGGAAAAATCGAA	1020
Db	321	LvsLysValGlnValSerLvsGluAsnThrThrIleIleAspGluValArgLysSerAla	340

1021	Qy	GATATTAAAGCTCGAGTTAAACAAATTCGTGCGCAAAATTCAGAGAAACAAGCTCAGATTAT	1080
341	Db	ThrIleGluAlaArgValGlyGlnIleLysThrGlnIleGluAspThrSerSerAspTyr	360
1081	Qy	GATCGTGAAAACTTCAAGAACGTCCTGTGCAAAACTTGTTCGTGGAGTAGCTGTTATCCAT	1140
361	Db	AspArgGlnLysIleuGlnIleuArgValAlaLysIleuAlaGlyGlyValAlaValIleLys	380
1141	Qy	GTTCGGAGCTCTACTGAAACTGAAATCAAAGAGAAGAGGATCGTGPAGAAGATGCTCTA	1200
381	Db	ValGlyAlaSerThrGlnIleGluMetLysGluLysLysAlaArgValGluAspAlaLeu	400
1201	Qy	AATGCAACAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTACTGCTTTGTC	1260
401	Db	HisAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyValAlaLeuVal	420
1261	Qy	CGTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACCTTGCTGCACTT	1320
421	Db	ArgAlaLeuValAlaValGlyAsnLeuThrGlyAlaAsnGluAspGlnThrHisGlyIle	440
1321	Qy	AATATCATCCGTCGTTCTTTGAAGACCTTTTACGTCAAAATTCGTGCAATCTGGGTAT	1380
441	Db	GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu	460
1381	Qy	GAAGGTCTATTCTGTAGAAAAAGTTTCGTGAACCAAAAGATGGTTTTCGATTAAATGCT	1440
461	Db	GluProSerValIleIleuAsnLysValLysGluGlyThrGlyAsnIleGlyTyrAsnAla	480
1441	Qy	GCATCAGAGAAATATGAAGACCTTATTAAAGCTGGTGTCAATGATCCATAAAAAGTTACA	1500
481	Db	AlaAsnGlyGluPheGlyAspMetValGluPheGlyIleLeuAspProThrLysValThr	500
1501	Qy	CGTATTGCATTACAAATGCAGCATCNGTAGCCCTCTTACTTCTAACTACAGATGCGCT	1560
501	Db	ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet	520
1561	Qy	ATTTCGTAAAAACCAAGACCTTAAAAAAGATATGCCT---ATGCCT---GGCGGTGGTATG	1614
521	Db	ValAlaAsp-----AlaProLysLysAspGluProAlaMetProAlaGlyGlyGlyMet	538
1615	Qy	GGTGGTATGGTGGTATG	1632
539	Db	GlyGlyMetGlyGlyMet	544

RESULT 8

```

US-10-369-493-16217
; Sequence 16217, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16217
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16217

Alignment Scores:
Pred. No.: 544
Score: 365
Percent Similarity: 81.87%
Conservative: 82

```

Best Local Similarity: 66.85% Mismatches: 95
 Query Match: 61.76% Indels: 4
 DB: 12 Gaps: 3

US-09-077-574A-1 (1-1647) x US-10-369-493-16217 (1-544)

QY 1 ATGGCTTCTAAAGAAATCCCTTTTGTGCTAAAGCCGCGTAAAGAACTTTCACAGAGGTGA 60
 DB 1 MetAlaAlaLysAspIleArgPheGlyGluAspAlaArgThrArgMetValArgGlyVal 20
 QY 61 GATAAACTTGCCTGCTTAAAGTAAACACTTGGACCTAAAGCCGCTAAATGCTGTTAT 120
 DB 21 AsnValLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40
 QY 121 GAAAGCTTTTGGTCCCGCAGTTATTTACAAAGATGCTGTATCTGTGCTGCTGCTGCT 180
 DB 41 GluLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60
 QY 181 GAACTTGAAGATAAGTTTGAATAATATCGCGCTCAAAATGTTTAAAGAAAGTAGCTCCCAA 240
 DB 61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80
 QY 241 ACTAGCGATATGCTGCTGATGGAACCTACACAGCAACAGCTCCCTGCGACAGCTATTAT 300
 DB 81 ThrAsnAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaLeuIle 100
 QY 301 CGTGAAGGTGTAACCTTGTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 DB 101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
 QY 361 GATAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 121 AspGlnAlaValLysAlaAlaValLysGluLeuLysAsnIleSerLysProThrThrAsp 140
 QY 421 CAAAAGAAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGATACCAACAACTAGGTAAT 480
 DB 141 AspLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160
 QY 481 ATCATAGCTGAAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 DB 161 IleIleAlaGluAlaMetGlnLysValGlyLysGluGlyValIleThrValGluGluGly 180
 QY 541 AAAGCTTCTGAACTACATAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 181 SerGlyLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyLysLeu 200
 QY 601 TCTCCATCTTGTAACTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 201 SerProTyrPheIleAsnAsnGlnGlnSerGlnSerAlaAspLeuAspAspProPheIle 220
 QY 661 CTTTGTATGAGAAAGATTTAGCATGATGAAAGACATGCTTACCAATCTTAGAACAGTT 720
 DB 221 LeuLeuHisAspLysLysIleSerAsnValArgAspLeuLeuProValLeuGluGlyVal 240
 QY 721 GCTAAAGTAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 DB 241 AlaLysAlaGlyLysProLeuLeuIleValAlaGluValGluGlyGluAlaLeuAla 260
 QY 781 ACATCTGTAGTCAATAGCTCCGCTGGAACCTCAAGTTGTAGCCGCTGCTGCTGCTGCTGCT 840
 DB 261 ThrLeuValValAsnThrIleArgGlyIleValLysValValAlaValLysAlaProGly 280
 QY 841 TTTGGTGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 DB 281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyThrVal 300
 QY 901 ATATTTGAAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 DB 301 IleSerGluGluValGlyLeuAlaLeuGluLysAlaThrIleLysAspLeuGlyArgAla 320
 QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 DB 321 LysLysValGlnValSerLysGluAsnThrThrIleIleAspGlyAlaGlyAspSerAla 340

QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAGAAACAAAGCTCAGATTAT 1080
 DB 341 ThrIleGluAlaArgValGlyGlnIleLysThrGlnIleGluAspThrSerSerAspTyr 360
 QY 1081 GATCGTGAAGAACTTCAAGAACTCTTGTCAAACTTGTGTGCTGCTGCTGCTGCTGCTGCT 1140
 DB 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
 QY 1141 GTTGGAGCTCTACTGAACTGAAATGAAAGAGAGAGAGAGAGATCGTGTAGAGATGCTCTA 1200
 DB 381 ValGlyAlaSerThrGluIleGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
 QY 1201 AATGCAACAGAGCTCGCTTGAAGAGATTTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 DB 401 HisAlaThrArgAlaAlaValGluGluGlyValProGlyGlyGlyValAlaLeuVal 420
 QY 1261 CGTCCATTAAAGCTCTGATGATATTAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 DB 421 ArgAlaLeuValAlaValGlyAsnLeuThrGlyAlaAsnGluAspGlnThrHisGlyIle 440
 QY 1321 AATATCATCGCTGCTCTTGTGAAGACCTTTAGTCAATTCCTCAAAATGCTGGCTAT 1380
 DB 441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460
 QY 1381 GAAGGTTCTATTGTTGTAGAAAGATTCGTGTAACCAAGATGCTTGTGATTTAATGCT 1440
 DB 461 GluProSerValIleLeuAsnLysValLysGluGlyThrGlyAsnIleGlyTyrAsnAla 480
 QY 1441 GCATCAGGAGATATGAAGACCTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 DB 481 AlaAsnGlyGluPheGlyAspMetValGluPheGlyIleLeuAspProThrLysValThr 500
 QY 1501 CGTATTGCATTCAAAATGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 DB 501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520
 QY 1561 ATTGCTGAAACCAAGACCTTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1614
 DB 521 ValAlaAsp-----AlaProLysLysAspGluProAlaMetProAlaGlyGlyMet 538
 QY 1615 GGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1632
 DB 539 GlyGlyMetGlyGlyMet 544

RESULT 9
 US-10-369-493-15265
 ; Sequence 15265, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15265
 ; LENGTH: 543
 ; TYPE: PRT
 ; ORGANISM: Agrobacterium tumefaciens
 US-10-369-493-15265

Alignment Scores:
 Pred. No.: 3,93e-141 Length: 543
 Score: 1828.50 Matches: 374
 Percent Similarity: 81.77% Conservative: 70

Best Local Similarity: 68.88% Mismatches: 96
Query Match: 61.61% Indels: 3
DB: 12 Gaps: 3

US-09-077-574a-1 (1-1647) x US-10-369-493-15265 (1-543)

```
QY 1 ATGGCTTCTAAAGAAATCCTTTTATGATGCTAAAGCCCGTGAAGAAATCTTTCAGCAGGTGA 60
DB 1 MetAlaAlaLysGluValLysProGlyAlaSerAlaArgGluLysMetLeuLysGlyVal 20
QY 61 GATAAAGTCTGAAATGCTTTAAAGTAAACACTTGGACCTAAAGCCGTAATGCTGTTAT 120
DB 21 AspIleLeuAlaAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40
QY 121 GAAAGTCTTTGGTTCCTCCAGTATTACAAAGATGGTGTATCTCTCCAAAGAAAT 180
DB 41 AspLysSerPheGlyAlaProPheIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACTTGAAGATAAGTTTGAAGATATGGCGCTCAAAATGGTTAAAGAACTAGTCCCAA 240
DB 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80
QY 241 ACTAGGATATGCTGCTGATGGAACACTACACAGCAACAGTCTCTGCAACAGTATTAT 300
DB 81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CGTGAAGGTGTAAGTCTGACAGCTGCTGCTTAATCTATGCCATTAAAGCTGCATA 360
DB 101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTGTTGTTGCTTACTAAAGAACTAAAGCGACATACAAAGCTACTCTGAC 420
DB 121 AspLeuAlaValAlaGluValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr 140
QY 421 CAAGAAGAAATGCTCAAGTCTGGAACCAATCTTCGAACTCTGATACAAATAGTAAT 480
DB 141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160
QY 481 ATCATAGCTGAAGCTATGCTAAAGTGAAGAGGAGGTGTTATCACAGTTCAGGAAGCT 540
DB 161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluGluAla 180
QY 541 AAAGTCTTGAACACTACATGATGTTGTAAGGAATGAAGTTTGACCGTGGCTACCTC 600
DB 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyLysLeu 200
QY 601 TCTCATATCTTGTAACTAATCTGAGAAATGGTTTGTGAACCTGATGAACCTTATATC 660
DB 201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220
QY 661 CTTTGTGAATGAGAAAGATTAAGTACATGATGTAAGCATGCTACCAATCTTAGACAAGTT 720
DB 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAGTAAACCGTCCACTCTCTATTATGCTGAAGAGCTAGAAGGTGAAGCACTTGA 780
DB 241 ValGlnThrGlyLysProLeuValIleAlaGluAspValGluGlyGluArgLeuAla 260
QY 781 ACACCTGTAGTCAATAGCTCGTGAGCACTCAAGTGTGTAGCCGTAAAGCTCTGTGT 840
DB 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280
QY 841 TTTGTGTGAAGCGCTGAAGCTATGCTGAAGATATTGCTATCTTACTGAGGAGAGCA 900
DB 281 PheGlyAspArgGlyAlaMetLeuGluAspIleAlaIleLeuThrGlyThrVal 300
QY 901 ATATTGAGATCGTGGTATAAGCTGAAATCTGTAAGCTGCTCTTCTTAGGAACAGCT 960
DB 301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyLysSer 320
QY 961 AAACGTGTAGTTATGCAAGAAATACTACTATCTGTTGATGGTGGTGAAGAAATCAGAA 1020
DB 321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyValaGlyGlnLysSer 340
```

```
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATTAT 1080
DB 341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluThrThrSerAspTyr 360
QY 1081 GATCGTGAAGAACTTCAAGACGCTCTTGCAGAACTTTGTTGGTGGAGTAGCTGTATCCAT 1140
DB 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleAsp 380
QY 1141 GTTGAGAGCTCTACTGAACTGAATGAAAGAGAGAGATCGTAGAAGATGCTCTA 1200
DB 381 ValGlySerThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGTATTGCTCCCTGGTGGTGTCTGCTTTGTC 1260
DB 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyValAlaLeuLeu 420
QY 1261 CCTCCATTAAAGTCTCTGATGATATTAAACCTGCTGATGATGATGATGATGATGAT 1320
DB 421 ArgSerSerThrLysIle--ThrValLysGlyValAsnAspAspGlnGluAlaGlyIle 439
QY 1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTTACGTCAATTTGCTCAATGCTGCTAT 1380
DB 440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459
QY 1381 GAAGTCTCTATTGTTGTAGAAAAGTTCTGTGAACCA--AAAGATGGTGTGATTTAAT 1437
DB 460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479
QY 1438 GCTGATCATGAGAAATATGAAGACCTTATTAAAGCTGTGTGTCATTCATGATCCTAAAAGTT 1497
DB 480 AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499
QY 1498 ACACGTATTGATTAACAAATCGACATCAGTAGCTCTTACTTCTTAACAGATGC 1557
DB 500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerValIleThrThrGluAla 519
QY 1558 GCTATTGCTGAAAACCAAGCACTTAAAGATATGCTCT--ATGCTGCGGTGGTATG 1614
DB 520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyMet 539
QY 1615 GGTGGTATG 1623
DB 540 GlyGlyMet 542
RESULT 10
US-10-369-493-11922
; Sequence 11922, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11922
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-11922
Alignment Scores: 6.3e-141 Length: 542
Score: 1826.00 Matches: 366
Percent Similarity: 82.59% Conservative: 80
```

Best Local Similarity: 67.78% Mismatches: 92
Query Match: 61.52% Indels: 2
DB: 12 Gaps: 2

US-09-077-574A-1 (1-1647) x US-10-369-493-11922 (1-542)

QY 1 ATGCTTCTAAAGAAATCCCTTTTGGCTAAAGCCGCTGAAAAAATCTTCCACGAGGTGA 60
DB 1 MetAlaAlaLysGluValLysPheHisSerAspAlaArgGluLysMetLeuArgGlyVal 20

QY 61 GATAAACTTGCACAAATGCTGTTAAAGTAACACTTGGACCTAAAGCCCGTAATGCTGTTATT 120
DB 21 AsnLeuLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValle 40

QY 121 GAAAGTCTTTGGTCCCGCATATTAACAAAGATGCTGTACTCTGTTGCAAAAGAAATT 180
DB 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluile 60

QY 181 GAACTTCAAGATAAGTTTGAATAATATCGCGCTCAAAATGGTTAAAGAAAGTAGCTCCCAAA 240
DB 61 GluLeuGluAspLysPheGluAsnMetGlyValAlaGlnMetValArgGluValAlaSerLys 80

QY 241 ACTAGCATATTCGTGTGATGTAACACACAGCAACAGTCCCTTGCAAGCTATTATTAT 300
DB 81 ThrSerAspLeuAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaValle 100

QY 301 CGTGAAGGTGTAACCTTGTAGAGCTGGTGGTAACTCTATGCGCATTAACGTTGGCATA 360
DB 101 LysGluGlyAlaLysAlaValAlaSerGlyMetAsnProMetAspLeuLysArgGlyile 120

QY 361 GATAAAGCTGTGTGTGTTTACTTAAAGAACTAAGCGACATTAACAAAGCCTACTCGTGAC 420
DB 121 AspLysAlaValGluAlaIleValGlnGluLeuLysThrAsnAlaArgLysValThrArg 140

QY 421 CAAAAAGAAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGATCAACAATAGGTAAT 480
DB 141 AsnAspGluileAlaGlnValGlyThrIleSerAlaAsnGlyAspAlaGluileGlyArg 160

QY 481 ATCATAGCTGAAGTATGGCTAAAGTTGGAAAGAGGTGTTATCAAGTTGAGGAAGCT 540
DB 161 PheLeuAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

QY 541 AAAGGTCTTGAACATACATTAAGTGTGTAAGGAATGAAGTTGACCGTGGCTACCTC 600
DB 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

QY 601 TCTCCATCTTTGTAACATCTCTGAGAAATGTTTGTGAACCTTGATAACCCCTTATATC 660
DB 201 SerProTyrPheIleThrAsnGlnAspLysMetArgValGluLeuGluProTyrVal 220

QY 661 CTTTGTATGAGAAAGATTACTAGCATGAAGACATGTACCAATCTTAGAACAAGTT 720
DB 221 LeuIleHisGluLysLysLeuSerAsnLeuGlnAlaLeuLeuProValLeuGluAlaVal 240

QY 721 GCTAAAGTAACCTCCACTCTTATTATTCCTGAAGACGTGAGGTGAAGCACTTGCA 780
DB 241 ValGlnSerLysLysLeuLeuIleLeuAlaGluAspValGluGlyGluAlaLeuAla 260

QY 781 ACACCTTGTAGTCAATAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCCCTGGT 840
DB 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysValAlaAlaValLysAlaProGly 280

QY 841 TTTGGTGAACCGCTAAAGCTATCTGTAAGATATTCCTATCTCTTCTTAGGACAGCT 900
DB 281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrAla 300

QY 901 ATATTTCAAGTCTGTGTATTAAGCTGAAATCTGAAATCTGCTTCTTAGGACAGCT 960
DB 301 IleSerGluAspLeuGlyIleLysLeuGluAsnValThrLeuGluMetLeuGlyArgAla 320

QY 961 AAACGCTGTAGTTATGCAAGAAATACTACTACTGTTGATGCTGTGCTGGAATAATCAAA 1020
DB 321 LysLysValValIleGluLysGluAsnThrThrIleValAspGlyAlaGlySerLysAsp 340

QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAGAAACAAAGCTCAGATTAT 1080
DB 341 GluileGlnGlyArgValSerGlnIleLysSerGlnIleGluThrThrSerAspTyr 360

QY 1081 GATCGTGAATAAATCTCAAGAACCTTCTTGAACAACTTGTGAGTAGCTGTTATCCAT 1140
DB 361 AspLysGluLysLeuGlnArgLeuAlaLysLeuAlaGlyValAlaValleArg 380

QY 1141 GTTGGAGCTCTACTGAACTGAAATGAAAGAGAGAGATCGTGTAGAGATGCTCTA 1200
DB 381 ValGlyGlySerThrGluValGluValLysGluArgLysAspArgValAspAlaMet 400

QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATGTCCTCGTGTCTGCTGCTCTTTGTC 1260
DB 401 HisAlaThrArgAlaAlaValGluGluGlyValLeuProGlyGlyValAlaLeuLeu 420

QY 1261 CGTCCATTAAGTCCCTTGTATGATTAATAACCTGCTGATGATGATGATGATGATGAT 1320
DB 421 ArgAlaAlaLysAlaLeuAspSerValGlnAlaGluAsnGluAspGlnLysHisGlyile 440

QY 1321 AATATCATCGTCTGCTTCTTGAAGCCCTTGTAGTCAAAATGCTGCAAAATGCTGGCTAT 1380
DB 441 GluileValArgArgAlaIleGluAlaProValArgGlnIleAlaGluAsnAlaGlyAla 460

QY 1381 GAAGTTCTATTCTGTAGAAAAGTTCCGTGAA---CCAAAAGATGCTTTTGGATTTAAT 1437
DB 461 GluGlySerIleIleValGlyLysLeuArgGluLysProGluPheGlyTyrPasn 480

QY 1438 GCTGCATCAGCAATATGAAGACCTTATTAAAGCTGCTGATGATGATGATGATGATGAT 1497
DB 481 AlaGlnThrAsnAlaPheGlyAspLeuTyrAsnGluGlyValIleAspProValLysVal 500

QY 1498 ACAGTATTCGATCAAAATGAGCATCAGTACGCTCTTACTTCTTAACACTAGAGATGC 1557
DB 501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrGluAla 520

QY 1558 GCTATTGCTCAAAACCCAGAACCTTAAAGATATGCT---ATGCTGCGGTGGTATG 1614
DB 521 MetValAlaGluLysProLysLysGluProAlaValProAlaMetProAlaGlyMet 540

RESULT 11
US-10-369-493-12330
; Sequence 12330, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12330
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12330

Alignment Scores:
Pred. No.: 8,08e-140 Length: 551
Score: 1812.50 Matches: 375
Best Local Similarity: 80.65% Conservative: 71
Query Match: 67.81% Mismatches: 96
Indels: 11
Gaps: 5

US-09-077-574A-1 (1-1647) x US-10-369-493-12330 (1-551)

QY	1	ATGCTTCTTAAAGAAATCCTTTTGTATGCTAAAGCCCGTGAAAAACCTTTCACGAGGTGA	60
Db	1	MetAlaAlaLysAspValLysPheSerArgAspAlaArgGluArgMetLeuArgGlyVal	20
QY	61	GATAAACCTTCCAATGCTGTAAAGTAACACTTGGACTTAAGGCGCTTAATGCTGTATT	120
Db	21	AsnLeuAlaLysAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValle	40
QY	121	GAAGAAGCTTTTGGTTCCCAAGTATTACAAAAGATGGTGATCTGTGTCAAAAGAAATT	180
Db	41	AspLysSerPheGlyAlaProArgGlyLeThrLysAspGlyValThrValAlaLysGluLe	60
QY	181	GAATCTGAAGATTAAGTTTGAATAATATGGCGCTCAAAATGGTTAAAGAAGTAGTCCCAA	240
Db	61	GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys	80
QY	241	ACTAGCGCATATTCGTCGTGATGGAACTACACACGACACAGTCCTTGCACAGCTATTAT	300
Db	81	ThrAsnSPileAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerileVal	100
QY	301	CGTGAAGCTGTAATAACTTTGACGAGCTGGTCGTAATCCTATGGCCATTAACACGTGGCATA	360
Db	101	GlnGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLe	120
QY	361	GATAAAGCTGTTGCTGCTTACTAAAGAACTTAAGCGACATTAACAAGCCTACTCTGTGAC	420
Db	121	AspLeuAlaValSerGluValValAlaAlaLeuGlyLysAlaAlaLysLysIleLysThr	140
QY	421	CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACCAAGATAGGTAAT	480
Db	141	SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyAspGluSerValGlyLys	160
QY	481	ATCATAGCTGAAGCTATGGCTAAAGTTGGAAGAGGAGTGTTATCACAGTTGAGGAAGCT	540
Db	161	MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla	180
QY	541	AAAGCTCTGAAGCTACATTAGATGTCGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC	600
Db	181	LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu	200
QY	601	TCCTCATCTTGTAACTAATCCTCGAAAAATGCTTGTGAACCTTGATACCCCTTATATC	660
Db	201	SerProLysPheValThrAsnAlaAspLysMetValAlaAspGluGluAspAlaTyrIle	220
QY	661	CTTTGTAAATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTT	720
Db	221	LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal	240
QY	721	GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGCTAGAGGTGAAGCACTTGCA	780
Db	241	ValGlnThrSerLysProLeuLeuIleSerGluAspValGluGlyGluAlaLeuAla	260
QY	781	ACACTTGTACTCAATTAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAGAAGCTCCGTGT	840
Db	261	ThrLeuValValAsnLysLeuArgGlyLysLysIleAlaValLysAlaProGly	280
QY	841	TTTGGTGAACCGCTAAAGCTATGCTTGAAGATATTCTATCTTACTGAGGAGAGACGA	900
Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal	300
QY	901	ATATTTCAGATCGTGGTATAAAGCTTGAAGATGAAGCTTGCTCTCTTTAGNACAGCT	960
Db	301	IleSerGluAspLeuGlyIleLysLeuGluAsnValGlyLeuAsnMetLeuGlyArgAla	320
QY	961	AAACGCTAGTATTGTACAAAGAAAAATCTACTATCTGCTGCTGTGCTGAAATACGAA	1020
Db	321	LysLysValSerIleSerLysGluAsnThrIleValAspGlyAlaGlyLysLysAla	340
QY	1021	GATATTAAAGCTCGAGTTAAACAAATTCGTGCGCAAAATTCAGAAAAACAGCTCAGATTAT	1080
Db	341	GluIleGlnGlyArgValAlaGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr	360

	Qy	1081	GATCGTGAAGAAACCTTCAGAACGCTTTGGCAAACTGTGGTGAGTAGCTGTATTCCAT	1144
	Dd	361	AspLysGluIlysLeuGlnArgLeuAlaLysLeuAlaGlyValAlaValIleArg	380
	Qy	1141	GTITGGAGCTCCTACTGAAACTGAAATCAAAGAAGGATCGTGTAGAAGATGCCTTA	1200
	Dd	381	valGLYglyAlaThrGluValGluValLysGluLysAspargValaspaAlaleu	400
	Qy	1201	AATGCCAACAAGAGCTCGGTTGAAGAAGGTATTGTCCCTGGTGGTGTACTGTTTTGTC	1260
	Dd	401	AsnAlaThrArgAlaalavalGluGluGlyIleValalagLyGlyValAlaleuLeu	420
	Qy	1261	CGCTCCCATTTAAAGTCCCTGGATGATATTAAACCT-----GCTGATGATGATGAACCTTGCT	1314
	Dd	421	ArgAlaSerAla-----AsnIleLysAlaalAgLYalaAsnalaspGlnAlaAla	437
	Qy	1315	GGACTTAAATATCATCCGTCGTTCTCTTTGAAGAGCCTTTACGTCAAAATTCGTGCAATGCT	1374
	Dd	438	GlyILEasnIleValargArgAlaleuGlnAlaproAlaarGInIlealaseraenAla	457
	Qy	1375	GGCTATCAGAGGTTCTATTGTGTGAGAAAAAGTTTCGTGAACCAAAAGATGCT---TTTGGA	1431
	Dd	458	GlyAlaGlualaserileValAlaGLYLysIleLeuGluasnLysGlyAlaThrPheGly	477
	Qy	1432	TTTTAANTGCTGCATCAGAGAGATATGAAGACCTTATTAAAGCTGGTGCATGTGATCCTAAA	1491
	Dd	478	TyrAsnAlaGlnThrGlyGluTyriGlyAspMetIleAlametyIleValaspProVal	497
	Qy	1492	AAAGTTACAGTATTTCATTACAAAATGCAGACATCAGTAGCCTCCCTACTCTTAACCTACA	1551
	Dd	498	LysValValargThrAlaleuGlnAspalaaLaservAladgyleuleuValthrThr	517
	Qy	1552	GAATGGCGCTATTGCTGAAATAACCAGAACCTCAAAAAA-----GATATGCGCT-----	1596
	Dd	518	GlualaMetIlealaglualaProLysLysGluSerAlaclyGlyGlyMetProGlyGly	537
	Qy	1597	ATGCCCTGGCGGTATGGGTGGTATGGGTGGGTATGGAC	1635
	Dd	538	MetcLYglvLYglvMetGlyMetGlyMetGlyVglwMetasp	550

RESULT 12

US-10-369-493-15838
; Sequence 15838, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-1052052/B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15838
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15838

Alignment Scores:
Pred. No.: 1,07e-139 Length: 540
Score: 1811.00 Matches: 361
Percent Similarity: 81.73% Conservative: 82
Best Local Similarity: 66.61% Mismatches: 95
Query Match: 61.02% Indels: 4
DB: 12 Gaps: 3

US-09-077-574A-1 (1-1647) x US-10-369-493-15838 (1-540)

Qy	1	ATGCGTCTCTAAAGAAATCTCTTTTATGCTGTAAAGCCCGTGAAGAAACTTTTCAAGAGGTGTA	60
Db	1	MetAlaAlaLysAspIleArgPheGlyGluAspAlaArgThrArgMetValArgGlyVal	20
Qy	61	GATAAACTTGCAATGCTGTGTAAAGTAAACACTTGGACCTCAAGGCGGTAAATCTCGTTATT	120
Db	21	AsnValLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu	40
Qy	121	GAAGAAGTCTTTTGGTTCCTCCAGTTATTACAAAGATGGTGTATCTGTGTGCAAAAGAAAT	180
Db	41	GluLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluLeu	60
Qy	181	GAACCTTGAAGATAAGTTTGAATAATGGCGCTCAAAATGGTTAAAGAAAGTAGTCCCAAA	240
Db	61	GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg	80
Qy	241	ACTAGCGATATGCTGGTGATGGAGTACACAGCAACAGTCTTGCAAGCTATTAT	300
Db	81	ThrAsnAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaLeuLeu	100
Qy	301	CGTGAAGGTGTAACACTTGTAGCAGCTGGTCTAATCTATGCCATTAAGCGTCGCATA	360
Db	101	ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLeu	120
Qy	361	GATAAAGCTGTGTGTCTTACTTAAGAACTTAAGCAGACATTACAAGCCTACTCGTGAC	420
Db	121	AspGlnAlaValLysAlaAlaValIleGluLeuLysAsnIleSerLysProThrThrAsp	140
Qy	421	CAAAAAGNAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACACAACTAGTAA	480
Db	141	AspLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn	160
Qy	481	ATCATAGCTGAAGCTATCGCTAAAGTTGCGAAAGGAGGTGTATCACAGTTCAGGAAGCT	540
Db	161	IleIleAlaGluAlaMetGlnLysValGlyLysGluGlyValIleThrValGluGlyLeu	180
Qy	541	AAAGTCTTGAACACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGCCTACCTC	600
Db	181	SerGlyLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu	200
Qy	601	TCTCCATCTTTGTAACTAACTCCTGAGAAATGGTTTGTGAACTTGATTAACCCCTTATATC	660
Db	201	SerProTyrPheIleAsnAsnGlnGlnSerGlnSerAlaAspLeuAspAspProPheIle	220
Qy	661	CTTTGTATGAGAAAAAGATTAGCATGAAAGACATGCTACCATTCTTAGACAAGTT	720
Db	221	LeuLeuHisAspLysLysIleSerAsnValArgAspLeuProValLeuGluGlyVal	240
Qy	721	GCTAAAGTAAACCGTCCACTCTTATTATGCTGAAGACGTAGAGAGTGAACACTTGCA	780
Db	241	AlaLysAlaGlyLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla	260
Qy	781	ACACTTGTAGTCAATAAGCTCGTGGAGACATCCCAAGTTGTAGCCGTAAAGCTCCTGGT	840
Db	261	ThrLeuValValAsnThrIleArgGlyIleValLysValValAlaValLysAlaProGly	280
Qy	841	TTTCGTGAACCGCTGAAGCTATGCTGAAGATATTGCTATCTTACTCGAGGAGAGCA	900
Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal	300
Qy	901	ATATTGTAAGATCGGTGTATAAGCTTGAAAATGTAAGCTTCTCTTTTGAACAACGCT	960
Db	301	IleSerGluGluValGlyLeuAlaLeuGluLysAlaThrIleLysAspLeuGlyArgAla	320
Qy	961	AAACGTGTAGTTATGCAAAAGAAATACTACTATCTGTGTATGCTGGTGGGAAATCAGAA	1020
Db	321	LysLysValGlnValSerLysLysAsnThrThrIleAspGlyAlaGlyAspSerAla	340
Qy	1021	GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATGGTGAAGAAACAGCTCAGATTAT	1080
Db	341	ThrIleGluAlaArgValGlnIleLysThrGlnIleGluAspThrSerSerAspTyr	360

Qy		1081	GATCGTGAATAAACTTCAGAACGCTCTTGCAAAACTCTGGTGTGAGTAGCTGTATTCCAT	1140
Db		361	AspargGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys	380
Qy		1141	GTTTGCAGCTGCTACTCGAATCGAAACGAGAGAAGGATCGGTAGAGAGATGCTCTTA	1200
Db		381	ValGlyAlaSerThrGluIleGluMetCysGlnLysLysalaArgValGluaspAlaLeu	400
Qy		1201	AATGCAACAAGAGCTGCGGTTTGAGAAGGTATTTCCTCGTGGTGGTACTGCTTTTTGTC	1260
Db		401	HisaLaThrArgAlaAlaValGluGluGlyValProGlyGlyGlyValAlaLeuVal	420
Qy		1261	CGTCCCATTAAGTCCTTGATGATATTAAACCTGCTGATGATGAATGAACCTTGTGGACTT	1320
Db		421	ArgAlaLeuValAlaValGlyAsnLeuThrGlyAlaAsnGluaspGlnThrHisGlyIle	440
Qy		1321	AATATCATCGTCCGTTCTTCTTGAAGAGCCTTTACGTCAAATGCTGCAAAATGCTGGCTAT	1380
Db		441	GlnIlealaLeuargAlaMetGluAlaproLeuargGlnIleValAlaasnAlaGlyGlu	460
Qy		1381	GAAGGTTCTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGTTTGGATTTTAATGCT	1440
Db		461	GluProSerValIleLeuAsnLysValLysGlnGlyThrGlyAsnTyrGlyTyAsnAla	480
Qy		1441	GCATCAGGAGATATGAAGACCTTATTAAAGCTGGTGTCAATTGATCCTAAAAGATTACA	1500
Db		481	AlaAsnGlyGluPheGlyAspMetValGluPheGlyIleLeuaspProThrLysValThr	500
Qy		1501	CGTATTTCATTACAAAATGCAGCATCAGTAGCCCTCTTCTTCAACTACAGATGCGCT	1560
Db		501	ArgSerAlaLeuGlnAsnAlaAlaserilGlaGlyLeuMetIleThrGluAlaMet	520
Qy		1561	ATTGCTGAAAAACCAGAACCTTAAAAAGATATGCT---ATGCCCT---GGCGGTGGTATG	1614
Db		521	ValAlaasp-----AlaProLysLysAspGluProAlaMetProAlaGlyGlyGlyMet	538
Qy		1615	GGTGGT	1620
Db		539	GlyGly	540

RESULT 13

```

US-10-369-493-12170
; Sequence 12170, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12170
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12170

```

Alignment Scores:

Alignment scores:	
Pred. No.:	1.3e-139
Score:	1810.00
Percent Similarity:	80.62%
Best Local Similarity:	67.57%
Query Match:	60.98%
DB:	12
Length:	552
Matches:	373
Conservative:	72
Mismatches:	99
Indels:	8
Gaps:	4

US-09-077-574A-1 (1-1647) x US-10-369-493-121170 (1-552)

```

QY 1 ATGGCTCTTAAGAAATCTTTTGTGATGCTAAAGCCGCTGAAACCTTTTACAGAGTGA 60
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 MetAlaAlaLysAspValLysPheSerArgAspAlaArgGluArgMetLeuArgGlyVal 20
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 GATAAATCTTGAATCTCTTAAAGTAACACACTTGGACCTTAAAGGCCGCTAATGCTTATT 120
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 GAAAGTCTTTTGGTCCCGATTAATCAAAAGATGGTGTATCTTGCAGAAAGAAAT 180
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle 60
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 GAACCTTGACATAGTTTGAATATGGCGCTCAATGTTAAAGAGTAGTCCCAAA 240
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 ACTAGCGATATTGCTGCTGATGAACTACAACAGCAACAGCTCTTGCACAAAGCTATTAT 300
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 CTGTAAGGTGTAACACTGTAGCAGCTGGTGTATCTTATGCCCATTAACGTCGATTA 360
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 101 GlnGluGlyHisLysAlaValAlaLagIleMetAsnProMetAspLeuLysArgGlyIle 120
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 361 GATAAAGCTGTTGCTCTTACTAAGAACTAAGACGACATATACAAAGCCTACTCTGAC 420
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 AspLeuAlaValSerAspValValThrLeuLysAsnAlaThrLysIleLysThr 140
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 421 CAAAAGAATAAGTCAAGTGAACCATTTCTGCAAACTCTGATACAAACATAGGTAAT 480
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 141 SerGluGluValAlaGlnValGlyThrIleAlaGlyAsnGlyAspGluSerValGlyLys 160
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGGAAGAGGTGTATATCAAGTTAGAGAGCT 540
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluAla 180
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 541 AAAGGCTTGAACATACATAGATGTTGAGGAATGAAGTTTTCACCGTGGCTACCTC 600
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyLysLeu 200
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 601 TCTCCATCTTTGTAATCTCTGAGAAATGTTGTGAACCTTGATCAACCTTATATC 660
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 201 SerProTyrPheValThrAsnAlaAspLysMetValAlaAspLeuGluAspAlaTyrIle 220
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 661 CTTTGTATGAGAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTT 720
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 721 GCTAAAGTAAACCGCTCACTCTTATTATTGCTGAAGACGTAGAAGTGAAGCACTGCA 780
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 ValGlnThrSerLysProLeuLeuIleSerGluAspValGluGlyGluAlaLeuAla 260
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 781 ACATTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTGCT 840
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaValLysAlaProGly 280
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 841 TTTGGTGAAGCCGTAAGCTATGCTTGAAGATATGCTTATCTTATCTGAGGAGAGCA 900
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 281 PheGlyAspArgGlyAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 901 ATATTTGAAGATCGTGTATAGCTTGAAGTGAAGTGAAGCTTGTCTTTTGAAGACAGCT 960
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 IleSerGluAspLeuGlyIleLysLeuGluAsnValGlyLeuAspMetLeuGlyArgAla 320
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 961 AAACGCTGATATTGACAAAGAAATATCTATCTATCTGATGCTGCTGAAATATCAGAA 1020
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyLysLysGlu 340
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAGCAACAGCTCAGATTAT 1080
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 341 GluIleGlnGlyArgValAlaGlnIleLysGlnIleGluGluThrThrSerAspTyr 360

```

```

QY 1081 GATCGTGAACAACTTCAAGAACGCTCTTGCAAACTTGTGTGGAGTAGCTGTATCCAT 1140
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 361 AspLysGluLysLeuGlnLargLeuAlaLysLeuAlaGlyValAlaValIleArg 380
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1141 GTTGAGAGCTGCTACTGAACTGAATGAAGAGAGAGAGATGCTGTAGAAGATGCTTA 1200
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 381 ValGlyGlyAlaThrGluValGluValLysGluLysAspArgValAspAlaLeu 400
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1201 AATGAACAGAGCTCGCTGAGAGAGATGCTGCTCCCTGGTGGTGTCTGCTTTGTC 1260
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 401 AsnAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyValAlaLeuLeu 420
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1261 CGCTCATTAAGTCTTCATGATATTAACCTGCTGTGATCATGATCAACTTGTGACTT 1320
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 421 ArgAlaSerLeuSerIleAsnAlaVal---GlyAlaAsnSerAspGlnThrAlaGlyIle 439
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1321 AATATCATCGCTGCTTCTTGAAGAGCCTTTACGTCAAAATGCTGCAATGCTGCTAT 1380
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 440 SerIleValArgAlaLeuGlnAlaProAlaArgGlnIleAlaAlaAsnAlaGlyAla 459
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1381 GAAGGTTCTTATGCTGAGAAAAGTTCGTGAACCAACCAAGATGCT---TTTGGATTAA 1437
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 460 GluAlaSerIleValAlaGlyLysIleLeuGluAsnLysGlyAlaThrPheGlyPheAsn 479
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1438 GCTGATCATGAGCAATATCAAGACCTTATTAAAGCTGCTGCTCATCTCTTAAAGATT 1497
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 480 AlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleValAspProValLysVal 499
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1498 ACACGATTGCTATTCAAAATGAGCATGAGTACCTCTTACTTCTTAATCAGAAATGC 1557
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 500 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla 519
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1558 GCTATGCTGAGAAACCAACCACTTAAACCACTTAAACCACTTAAACCACTTAAAC 1599
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 520 MetIleAlaGluAlaProLysLysGluSerAlaGlyGlyGlyMetProGlyGlyMet 539
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1600 CTTGGCGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1635
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 540 GlyGlyGlyGlyMetGlyGlyMetGlyGlyMetGlyGlyMetGlyGlyMetGlyGlyMet 551

```

RESULT 14

US-10-369-493-167
Sequence 167, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 167
LENGTH: 545
TYPE: PRT
ORGANISM: Aquifex aeolicus
US-10-369-493-167

Alignment Scores:
Pred. No.: 1.19e-137 Length: 545
Score: 1786.00 Matches: 352
Percent Similarity: 82.29% Conservative: 94
Best Local Similarity: 64.94% Mismatches: 86
Query Match: 60.18% Indels: 10
DB: 12 Gaps: 2

US-09-077-574A-1 (1-1647) x US-10-369-493-167 (1-545)

Qy	1	ATGCCTCTAAAGAAATCCTTTTGTATGCTTAATGAACCCCGTGAATAAATTTCACGAGGTGTA	60
Db	1	MetAlaAlaLysAlaIleIlefyAsnGluGluAlaArgAlaLysLeuLysAlaGlyVal	20
Qy	61	GATAAACTTGCAATGCTGTAAAGTAACACTTGACCTTAAGGCCGTAAATGTCCTTATT	120
Db	21	AspLysLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgGluValIleLeu	40
Qy	121	GAAGAAGTCTTTGGTTCCTCCAGTTATTACAAGAAGATGCTGATCTGTGTGCAAAAATAAT	180
Db	41	GlyLysAsnTrpGlyThrProValvalthrLysaspGlyValThrValalaysGluIle	60
Qy	181	GAAC TTGAAGATAAGTTTAAAAATATGGCGCTCAAATGGTTAAAGAAAGTAGTCCCCAAA	240
Db	61	GluLeuLysAspLysPheGluAsnIleGlyAlaGlnLeuValLysGluValAlaSerLys	80
Qy	241	ACTAGCATATTCGTGTGATGGAACACTAACAGCAACAGTCCTTGCCAAGCTATTATAT	300
Db	81	ThrAlaaspValAlaGlyAspGlyThrThrAlatrValLeuAlaGlnAlaIlePhe	100
Qy	301	CGTCAAGGTGTAATAACTTGTAGCACCTGTCGTATCTATGGCCATTAAACCGTGCSCATA	360
Db	101	HisGluGlyLeuArgvalAlaAlaSerGlyAlaasnValMetGluValLysargGlyIle	120
Qy	361	GATAAAGCTGTGTGCTGTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC	420
Db	121	AspLysAlaValLysLysIleValGluGluLeuLysLysLeuSerLysaspValLysGlu	140
Qy	421	CAAAAAGNAATAGTC AAGTTGG AACCAATTTCTGCAAACTCTGATACAACTAGGTAAAT	480
Db	141	ArgLysGluIleGluGlnValAlaThrIleSerAlaasnAsnAspProGluIleGlyLys	160
Qy	481	ATCATAGCTGAAGCTATCGCTAAAGTTCGAAAGGAGGTGTTATCACAGTTGAGGAAGCT	540
Db	161	IleIleAlaaspAlaMetGluGluValGlyLysAspGlyValIlethrValGluGluSer	180
Qy	541	AAAGGCTCTTGAACATACATTAGATGTGTTGAAGGAATCAAGTTTGACCGTGCTACCTC	600
Db	181	LysSerAlaGluThrThrLeuGluValVallysGlyMetGlnPheAspArgGlyIyrLeu	200
Qy	601	TCTCCATAC TTTGTA ACTAATCCTGAGAAAATGGTTTGTGAAC TTGATAAC CTTATATFC	660
Db	201	SerProtyrPheValThrAspProLunlysMetGluCysValLeuGluAsnProtyrIle	220
Qy	661	CTTTGTANTGAGAAAAGATTACTAGCATGAAGACATGCTACCATCTTAGAACAAAGTT	720
Db	221	LeulleTyrgluLysLysIlethrAsnVallysGluLeuLeuProIleLeuGluGlnVal	240
Qy	721	GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA	780
Db	241	ValArgserGlyArgProLeuLeuValIleAlaGluaspValGluGlyGluAlaLeuAla	260
Qy	781	ACACTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAAGCTCCTCGT	840
Db	261	ThrLeuValValAsnHisIleLysglyValLeuLysAlaCysAlaValLysAlaProgly	280
Qy	841	TTTTGGTGAACCCGTAAAGCTATGCTTGAAGATATTGCTATTCCTTACTGGAGGAGAACA	900
Db	281	PheGlyGlnArgArgLysAspTyrlauGlyAspIleAlaLeuThrGlyGlyGlnAla	300
Qy	901	ATATTTGAAGATCGTGGTATAAAGCTCAAAATGTAAGCTGTCTCTTTAGGACACAGCT	960
Db	301	IlethrGluaspLeuGlyIleLysleuGluSerValthrLeuAspMetLeuGlyGlnAla	320
Qy	961	AAACGCTAGTATTGACAAAGAAAATACTACTATPCGTTGATGGTGGCTGCAAAAATCAGAA	1020
Db	321	GluLysValValValAspLysGluHsfthrThrIleIleGlyGlyLysGlyAspProglu	340
Qy	1021	GATATTTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAAACAAGCTCAGATTAT	1080
Db	341	GlnIleLysAlaArqIleGlnIleLysArgGlnIleGlnIleGlnIleThrSerAspIvr	360

1081 GATCGTGAATAAATTCAAGAACGTCCTGCAAACTTCTTGGTAGAGTAGCTGTATCCAT 114
Db AspArgGluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaIlelleArg 380
1141 GTTGGAGCTCTACTGAACCTGAAATCAAGAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
Db ValGlyAlaAlaThrGluAlaGluLeuLysGluLysLysTyrArgValGluAspAlaVal 400
1201 AATGCACAAAGAGCTGCGGTTGAAGAAGGTATTGTCTCCCTGGTGTGCTACTGCTTTTGT 1260
Db HisAlaThrLysAlaAlaValGluGluGlyLeuValProGlyGlyGlyValAlaLeuVal 420
1261 CGCTCCAAATAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACCTTGTGGACTT 1320
Db ArgAlaSerGluAlaLeuGluAspLeuLysGlyAspAsnHisAspGlnGlnLeuGlyLe 440
1321 AATATCATCCGTCGTCCTCTTGAAGACCCCTTACGTCAAATTCGTGCAAAATGCTGGCTAT 1380
Db AspIlelleLysLysAlaValArgThrProLeuLysGlnIleAlaTyrAsnAlaGlyTyr 460
1381 GAAGGTTCTATTGTTGTAGAAAAAGTT-----CGTGAACCAAAAGATGGTTTT 1428
Db AspGlySerValValLeuGluLysValIleGluLeuLysGlyLysGlyValSerTyr 480
1429 GGATTTAATGCTGCATCAGAGAAATATGAAGACCTTATTAAAGCTGGTGTCATGTGACT 1488
Db GlyPheAsnAlaAlaThrGlyGlyTyrValAspMetTyrGluAlaGlyIlelleAspPro 500
1489 AAAAAAGTTACAGCTATTGCTATTACAAAAATGCAGCATCAGTACGCTCTTACTTCTAACT 1548
Db ThrLysValValArgThrAlaIleGluAsnAlaAlaSerValAlaGlyThrMetLeuThr 520
1549 ACAGATGCGCTATTGCTGAAAAAACCCAGAACCTTAAAAA-----GAT 1590
Db AlaGluAlaLeuIleAlaAspLeuProGluGluLysLysAspIleThrProThrAsp 540
1591 ATGCGCT 1596
Db MetPro 542

RESULT 15
US-10-369-493-17132
; Sequence 17132, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17132
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17132

Alignment Scores:		
Pred. No.:	3.37e-137	544
Score:	1780.50	355
Percent Similarity:	81.02%	Conservative: 89
Best Local Similarity:	64.78%	Mismatches: 95
Query Match:	59.99%	Indels: 9
DB:	12	Gaps: 4

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:23:01 : Search time 19.5 Seconds

(without alignments)
7147.280 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 2968

Sequence: 1 atggcttctaagaatacct.....gtatggacggtatgactag 1647

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p_model -DEV=xlip
-Q=/cgn2_1/USPRO_spool_p/US09077574/runat_28012004_125847_11658/app_query.fasta_1.1799
-DB=issued Patents AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=ext -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574.@CGN_1_1_27.@runat_28012004_125847_11658 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*

2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*

3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*

4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*

5: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pcp.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1844	62.1	548	2	US-08-467-822-31
2	1844	62.1	548	3	US-08-432-697-31
3	1844	62.1	548	3	US-08-466-248-31
4	1820.5	61.3	551	4	US-09-252-991A-24209
5	1768.5	59.6	550	4	US-08-328-352-5508
6	1741	58.7	548	2	US-08-467-822-32
7	1741	58.7	548	3	US-09-472-971-3
8	1741	58.7	548	3	US-08-432-697-32
9	1741	58.7	548	3	US-08-466-248-32
10	1723.5	58.1	546	3	US-08-470-260-6
11	1723.5	58.1	546	3	US-08-471-491-6
12	1723.5	58.1	546	3	US-08-466-662-6

13	1723.5	58.1	546	4	US-08-256-847C-1
14	1723.5	58.1	546	4	US-08-256-847C-7
15	1721	58.0	544	2	US-08-467-822-33
16	1721	58.0	544	3	US-08-432-697-33
17	1721	58.0	544	3	US-08-466-248-33
18	1716.5	57.8	547	4	US-08-461-722-2
19	1716.5	57.8	547	4	US-08-336-251-2
20	1716.5	57.8	547	4	US-09-468-041-2
21	1716.5	57.8	547	5	PCT-US94-06362-2
22	1708	57.5	544	4	US-09-198-452A-153
23	1702	57.3	545	2	US-08-467-822-30
24	1702	57.3	545	3	US-08-432-697-30
25	1702	57.3	545	3	US-08-466-248-30
26	1693.5	56.7	641	4	US-09-613-303-51
27	1597	53.8	541	2	US-08-997-080-160
28	1597	53.8	541	2	US-08-997-362-160
29	1597	53.8	541	3	US-09-095-855-160
30	1597	53.8	541	4	US-09-324-542-160
31	1597	53.8	541	4	US-09-205-426-160
32	1595.5	53.8	540	4	US-09-613-303-4
33	1595.5	53.8	639	4	US-09-613-303-17
34	1595.5	53.8	648	4	US-09-613-303-29
35	1595.5	53.8	948	4	US-09-613-303-21
36	1591.5	53.6	540	4	US-08-461-722-4
37	1591.5	53.6	540	4	US-08-336-251-4
38	1591.5	53.6	540	4	US-09-468-041-4
39	1591.5	53.6	540	5	PCT-US94-06362-4
40	1582.5	53.3	540	4	US-09-134-001C-4705
41	1577.5	53.2	541	2	US-08-467-822-34
42	1577.5	53.2	541	2	US-08-447-154-19
43	1577.5	53.2	541	3	US-08-432-697-34
44	1577.5	53.2	541	3	US-08-466-248-34
45	1575	53.1	540	2	US-08-368-834-20

ALIGNMENTS

RESULT 1

US-08-467-822-31
: Sequence 31, Application US/08467822
: Patent No. 5843460

: GENERAL INFORMATION:

: APPLICANT: Labigne, Agnes

: APPLICANT: Sauerbaum, Sebastien

: APPLICANT: Ferrero, Richard I.

: TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

: TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

: TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

: TITLE OF INVENTION: POLYPEPTIDES

: NUMBER OF SEQUENCES: 44

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

: ADDRESSEE: Dunner

: STREET: 1300 I Street, N.W.

: CITY: Washington

: STATE: D.C.

: COUNTRY: USA

: ZIP: 20005-3315

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent In Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/467,822

: FILING DATE: 06-JUN-1995

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/447,177

: FILING DATE: 19-MAY-1995

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

Db 462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481
QY 1447 GGAGAAATGAAGACCTTAAAGCTGGTGCATGTATCTTAAAGAAAGTTACAGTATT 1506
Db 482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501
QY 1507 GCATTACAAATGACGATCAGTACGCTTCTTACTTCTAACTACAGAAATGCGCTATTGCT 1566
Db 502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521
QY 1567 GAAACACAGAACCTTAAAGAAAGATATGCTATGCTGGCGGTGGTATGGGTGGTATGGGT 1626
Db 522 AspLeu-----ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539
QY 1627 GGTATGGAGCGGTATG 1641
Db 540 GlyMetGlyGlyMet 544

RESULT 3

US-08-466-248-31
Sequence 31, Application US/08466248
Patent No. 6258359
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thibierge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,248
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-248-31

Alignment Scores: 5,71e-170 Length: 548
Pred. No.: 1844.00 Matches: 366
Score: 82.02% Conservative: 81
Percent Similarity: 67.16% Mismatches: 96
Best Local Similarity: 62.13% Indels: 2
Query Match: 3 Gaps: 1
DB: US-09-077-574A-1 (1-1647) x US-08-466-248-31 (1-548)
QY 7 TCTAAGAAATCTTTTTCATGCTTAAGCCCTGAAAACTTTACAGAGGTGTAGATAAA 66
Db 2 AlaLysGluLeuArgPheGlyAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 21
QY 67 CTTGCAAAATGCTTTAAAGTAAACACTTGACCTTAAAGCCGCTAAATGCTTATTTGAAAAAG 126
Db 22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41
QY 127 TCTTTTGGTCCCAAGTATTACAAAAGATGGTGTATCTTGTTCGCAAGAAATGAACTT 186
Db 42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61
QY 187 GAAGATAAGTTTCAAAATATGGCGCTCAAAATGGTTAAAGAGCTAGCTCCCAAAATAGC 246
Db 62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81
QY 247 GATATTGCTGGTATGGAACCTACAAACAGCAACAGTCTCTTGCAAGCTATTTATCGTGA 306
Db 82 AspThrAlaGlyAspGlyThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101
QY 307 GGTGTAAGAACTTGTAGCAGCTGCTGTAATCTTATGCGCATTAAGCTGGCATAGATAAA 366
Db 102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121
QY 367 GCTGTGTGCTCTTACTTAAGAACTAAGCGACATTACAAAGCTACTCTGTCACCAAAA 426
Db 122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141
QY 427 GAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGATAACAATAGTAATATATCATA 486
Db 142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleIle 161
QY 487 GCTGAAGCTATGCTAAAGTTGGAAGAGGTGTTATACAGTTGAGAGCTAAAGGT 546
Db 162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181
QY 547 CTTGAAACTACATTAGATGTGTTGAGGATGAAGCTTTCACCGTGCCTCTCTCCA 606
Db 182 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201
QY 607 TACTTTGTAATCTCTGAGAAAATGGTTTGTGAACCTTGATACCCCTTATATCTTTGT 666
Db 202 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeu 221
QY 667 AATGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTTGCTAAA 726
Db 222 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241
QY 727 GTAACCGTCCACTCTCTTATTTGCTGAAGCTAGAGCTGAAGCTGAAGCACTTGAACACT 786
Db 242 SerGlyArgProLeuLeuIleAlaGluAspIleValLysValCysAlaValLysAlaProGlyPheGly 261
QY 787 GTAGTCAATAAGCTCCCTGGAGCACTCCCAAGTTGTAGCCGTAAAGCTCTCTGTTTGT 846
Db 262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281
QY 847 GAACCGCTTAAAGCTATGCTTGAAGATTTGCTTATCTCTTACTGAGGAGAGCAATATT 906
Db 282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 301
QY 907 GAAGATCGTGTATTAAGCTTGAATCTTAAGCTTGTCTTCTTTAGGAAACAGCTAAAGCT 966
Db 302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321

QY 967 GTAGTATTGACAAAGAAATCTACTATCTGTGCTGCTGGAATATCAGAGATATT 1026
Db : : : : :
QY 322 ILeValThrLysGluAsnThrThrIleLeuAspGlyGluGlyAlaThrGluLeu 341
Db : : : : :
QY 1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db : : : : :
QY 342 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 361
QY 1087 GAAGAACTTCAGAGACGCTTCGAAACTTGTGGTGGAGTAGCTGTATCCATGTTGGA 1146
Db : : : : :
QY 362 GluLysLeuGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381
QY 1147 GCTGCTACTGAACTGAAATGAAAGAGAGAGGATCGTGTAGAGATGCTCTAAATGCA 1206
Db : : : : :
QY 382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401
QY 1207 ACAAGAGCTGGTGTGAAGAGGATTTGTCCTCGTGTGCTACTGCTTTGTCCTGCTCC 1266
Db : : : : :
QY 402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyValAlaLeuIleArgAla 421
QY 1267 ATTAAGCTCTGTGATGATATAAACCCTGCTGATGATGATGATGATGATGATGATGAT 1326
Db : : : : :
QY 422 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 441
QY 1327 ATCCGCTGCTTCTCTGAAGAGCCCTTTACGTCAAATTCGTCAAATGCTGCTGAAGGT 1386
Db : : : : :
QY 442 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 461
QY 1387 TCTATTGCTGTGAGAAAAGTTCGTGAACCAAGATGTTTGGATTTAATGCTGCATCA 1446
Db : : : : :
QY 462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481
QY 1447 GGCAATATGAGACCTTATTAAGCTGGCTCAATGATGATGATGATGATGATGATGATGAT 1506
Db : : : : :
QY 482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501
QY 1507 GCATTACAAATGAGCATGAGTACCTCTTACTTCTTACTACAGATGCTGCTATGCT 1566
Db : : : : :
QY 502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521
QY 1567 GAAACACGAGACCTTAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1626
Db : : : : :
QY 522 AspLeu-----ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539
QY 1627 GGTATGACCGTATG 1641
Db 540 GlyMetGlyGlyMet 544

RESULT 4

US-09-252-991A-24209
; Sequence 24209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24209
; LENGTH: 551
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24209
Alignment Scores:
Pred. No.: 1.08e-167 Length: 551

Score: 1820.50 Matches: 366
Percent Similarity: 80.33% Conservative: 75
Best Local Similarity: 66.67% Mismatches: 103
Query Match: 61.34% Indels: 5
DB: 4 Gaps: 2
US-09-077-574A-1 (1-1647) x US-09-252-991A-24209 (1-551)
QY 1 ATGCTCTTAAAGAAATCCCTTTTGTAGTGTAAAGCCGTGAAATACTTTCACGAGGTGTA 60
Db : : : : :
QY 5 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysMetLeuValGlyVal 24
QY 61 GATAAATCTGCAATGCTGTTTAAAGTAACACTTCGACTGAAGCCGTAAGTGTGTTATT 120
Db : : : : :
QY 25 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValLeu 44
QY 121 GAAAGCTCTTTTGTGTTCCCGAGTTATTACAAAGATGCTGTATCTGTTGCAAAAGAAAT 180
Db : : : : :
QY 45 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 64
QY 181 GAATCTGAAGATAAGTTTGAATAATATGGCGCTCAAAATGTTTAAAGAGTAGTCCCAAA 240
Db : : : : :
QY 65 GluLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 84
QY 241 ACTAGCGATATTGCTGTGATGGAATCTACACAGCAACAGCTCTTGCACAGCTATTAT 300
Db : : : : :
QY 85 AlaAsnAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 104
QY 301 CGTGAAAGCTGTAAACCTTGTAGCAGCTGGTCTGTAATCTTATGGCCATTAAACCTGGCATA 360
Db : : : : :
QY 105 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 124
QY 361 GATAAAGCTGTTGCTGTGTTACTAAAGAACTAAGCAGCATTAACAAGCTTACTCTGAC 420
Db : : : : :
QY 125 AspLysAlaThrValAlaIleValAlaGlnLeuLysGluLeuAlaLysProCysAlaAsp 144
QY 421 CAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAACCTCTGATACACAACTAGGTAAT 480
Db : : : : :
QY 145 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyGln 164
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGGAAGAGGCTGTATCATCAGTTGAGGAAAGCT 540
Db : : : : :
QY 165 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 184
QY 541 AAAGCTCTGAAACTACATTAGATGTTGTTGAAGAAATGAAGTTGACCTGCTACTCTC 600
Db : : : : :
QY 185 SerGlyLeuGluAsnGluSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 204
QY 601 TCTCCATCTTTGTAATCTCTGAGAAATGTTTGTGAACTTGTGATACCTTATATC 660
Db : : : : :
QY 205 SerProTyrPheValAsnLysProAspThrMetAlaAlaGluLeuAspSerProLeuLeu 224
QY 661 CTTTGTAAATGAGAAAGATTTACTAGCATGAAAGACATGTCTACCAATCTTAGAACAAAGTT 720
Db : : : : :
QY 225 LeuLeuValAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 244
QY 721 GCTAAAGTAAACCGTCCCTCTCTTATTATGCTGAAGACCTGAGAGTGAAGCACTTGCA 780
Db : : : : :
QY 245 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 264
QY 781 ACATTGTAGTCAATAAGCTCCCTGGAGCACTCCAAGTTCTAGCCGTAAAGCTCCTGCT 840
Db : : : : :
QY 265 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 284
QY 841 TTTGCTGAACCGCTGAAAGCTATGCTTGAAGATATTGCTTATCTTACTGAGGAGAGCA 900
Db : : : : :
QY 285 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyThrVal 304
QY 901 ATATTTGAAGATCGTGGTATAAAGCTGTGAAATGTAAGCTGTGCTTTCTTTAGGAACAGCT 960
Db : : : : :
QY 305 IleSerGluGluValGlyLeuSerLeuGluGlyAlaThrLeuGluHisLeuGlyAsnAla 324
QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCTGTTGATGCTGCTGAAATACGAA 1020

Db 325 LysArgValValIleAsnLysGluAsnThrThrIleAspGlyAlaGlyValGlnAla 344
QY 1021 GATATTAAGCTCGAGTTAAACAAATCTGTGCACAAATGAAGAACAGCTCAGATTAT 1080
Db 345 AspIleGluAlaArgValLeuGlnIleArgLysGlnIleGluGluThrThrSerAspTyr 364
QY 1081 GATCTGGAACACTTCAAGACGCTCTGCAAACTTGTGTGGAGTAGCTGTATCCAT 1140
Db 365 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleLys 384
QY 1141 GTTGAGCTGCTACTCAAACTGAAATGAAGAGAGAGAGATCGTGTAGAGATCGCTTA 1200
Db 385 ValGlyAlaAlaThrGluValGluMetLysGluAsnLysAlaArgValGluAspAlaLeu 404
QY 1201 AATGCAACAGAGCTCGGTGAAGAGTATTGTCCTGCTGGTGGTACTGCTTTGTC 1260
Db 405 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 424
QY 1261 CGCTCCATTAAAGTCCTTCATGATATTAAACCTGCTGATGATGATGATGATGATGAT 1320
Db 425 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluGluGlnAsnValGlyIle 444
QY 1321 AATATCATCCGCTGCTCTGGAAGAGCTTTAGCTCAAAATTCGTCAAAATGCTGCTAT 1380
Db 445 AlaLeuLeuArgArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyAsp 464
QY 1381 GAAGGTTCTATTGTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTGGATTAAATGCT 1440
Db 465 GluProSerValValAspLysValLysGlnGlySerGlyAsnTyrGlyPheAsnAla 484
QY 1441 GCATCAGGAGATATCAAGACCTTATTAAAGCTGGTGTCTATGATCCTTAAAGAGTTACA 1500
Db 485 AlaThrGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 504
QY 1501 CGTATTGCAATTAACAAATCAGCATCAGTAGCCTCTTACTTCTTAATCAAGATCGCT 1560
Db 505 ArgSerAlaLeuGlnAlaAlaSerIleGlyLeuMetIleThrThrGluAlaMet 524
QY 1561 ATTGCTGAAAAACCAAGACCTTAAAGATATGCTTATGCTGCTGCTGCTGCTGCTGCT 1614
Db 525 ValAlaGluIleValGlu-----AspLysProAlaMetGlyGlyMetProAspMet 541
QY 1615 GTGGTATGGTGGTATGACGGTATG 1641
Db 542 GlyGlyMetGlyGlyMetGlyGlyMet 550

RESULT 5

US-09-328-352-5508
; Sequence 5508, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5508
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5508

Alignment Scores:
Pred. No.: 1,17e-162 Length: 550
Score: 1768.50 Matches: 348
Percent Similarity: 79.96% Conservative: 87
Best Local Similarity: 63.97% Mismatches: 108
Query Match: 59.59% Indels: 1
DB: 4 Gaps: 1

US-09-077-574A-1 (1-1647) x US-09-328-352-5508 (1-550)
QY 1 ATGGCTTCTAAAGAAATCTTTTTCATGCTAAAGCCGCGTAAAGAAATCTTTCACGAGGTGA 60
Db 7 MetSerAlaLysAspValLysPheGlyAspSerAlaArgSerLysMetIleAlaGlyVal 26
QY 61 GATAAATCTGCAATCTGCTTAAAGTAAACCTTGAACCTTAAAGCGCGTAAATCTGCTTAT 120
Db 27 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 46
QY 121 GAAAGTCTTTTGGTCTCCCACTATTACAAAAGATGCTGCTATCTGCTCAAAAGAAAT 180
Db 47 AspArgSerPheGlyAlaProHisIleThrLysAspGlyValThrValAlaLysGluIle 66
QY 181 GAACTTGAAGATAAGTTTGAAATATGGCGCTCAAAATGCTTAAAGAACTAGCTCCAAA 240
Db 67 SerLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValSerSerLys 86
QY 241 ACTAGCGATATTGCTGGTGTGATGGAACCTTACAAACAGCAACAGTCTCTTGCACAGCTATTAT 300
Db 87 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleLeu 106
QY 301 CGTGAAGGTGTAACCTTGTACAGCTGCTGTATCTTATGCTTATGCTTATGCTTATGCTTAT 360
Db 107 AsnGluGlyIleLysSerValThrAlaGlyMetAsnProMetAspLeuLysArgGlyIle 126
QY 361 GATAAAGCTGTTGTTGCTGTTACTTAAAGAACTAAGCGACATTAACAAAGCTACTCTCGTAC 420
Db 127 AspIleAlaValLysThrValGluAsnIleArgSerIleAlaLysProAlaAspASP 146
QY 421 CAAAAGAAATAGCTCAAGTTCGAACCTTCTGCAAACTCTGATACACATAGTAGTAT 480
Db 147 PheLysAlaIleGluValGlySerIleSerAlaAsnSerAspThrThrValGlyLys 166
QY 481 ATCATAGCTCAAGCTATGCTTAAAGTTCGAAAGAGGTGTTATCACAGTTCAGGAGCT 540
Db 167 LeuIleAlaGlnAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 186
QY 541 AAGGCTTGAACCTACATTAGTGTGTTGAAGAACTGAAGTTCACCGCTGCTGCTACCTC 600
Db 187 SerGlyPheGluAspAlaLeuAspValValGluGlyMetGlnPheAspArgGlyTyrIle 206
QY 601 TCTCCATCTTTGTAATCTCTGAGAAATAGTGTGTAAGTTCGAACTTGAACCTTATATC 660
Db 207 SerProTyrPheAlaAsnLysGlnAspThrLeuThrAlaGluLeuGluAsnProPheIle 226
QY 661 CTTTGTATAGCAAAAGATTTACTAGCATGAAGACATGCTACCATTCTTAGACAAAGTT 720
Db 227 LeuLeuValAspLysLysIleSerAsnIleArgGluLeuIleSerValLeuGluAlaVal 246
QY 721 GCTAAAGTAAACCGTCCCTCTTATTATTGCTGAAGAGCTAGAAGTGAAGCACTTGCA 780
Db 247 AlaLysThrGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 266
QY 781 ACATTTGTAGTCAATTAAGTCCGTGAGCAGCTCCAAAGTGTAGCCGTAAAGCTCTGGT 840
Db 267 ThrLeuValValAsnAsnMetArgGlyIleIleLysValCysAlaValLysAlaProGly 286
QY 841 TTTGCTGAACCGCTGAAGCTATGCTTGAAGATATTGCTTATCTCTTCTGAGGAGCA 900
Db 287 PheGlyAspArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyAlaThrVal 306
QY 901 ATATTTGAAGATCGTGTATTAAGCTTGAAGATTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db 307 IleSerGluGluValGlyMetSerLeuGluGlnAlaThrLeuGlnAspLeuGlyThrAla 326
QY 961 AAACGTGTAGTTATTGACAAAGAAATATCTATCTGTTGATGCTGCTGCTGCTGCTGCTGCT 1020
Db 327 HisLysIleThrValSerLysGluAsnThrValIleValAspGlyAlaGlyAspAlaAla 346
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAGCAAACTCAGATTAT 1080
Db 347 AlaIleAlaGluArgValGlnGlnIleArgAlaGlnIleGluSerThrSerGluTyr 366

```
QY 1081 GATCGTGAAGAACTTCAAGACGCTTGGCAAACTTGTGGTGGAGTAGCTGTTATCCAT 1140
Db 367 AspArgGluLysLeuGlnGluValAlaLysLeuAlaGlyGlyValAlaValLys 386
QY 1141 GTTGGAGCTGCTACTGAACATGAATGAAGAGAGAGAGATCGTGTAGAGATGCTCTA 1200
Db 387 IleGlyAlaAlaThrGluValGluMetLysGluLysAspArgValAspAlaLeu 406
QY 1201 AATCAACAGAGCTGGTGTGAAGAGGATTTGCTCCCTGGTGGTACTGCTTTGTC 1260
Db 407 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuVal 426
QY 1261 CGTCCATTAAGTCCTGATGATTAACCTGCTGATGATGATGAACCTTGTGGACTT 1320
Db 427 ArgAlaValAsnAlaLeuGluGlyLeuLysGlyAlaAsnGluAspGlnThrAlaGlyLe 446
QY 1321 AATATCATCGCTGCTTCTTGAAGAGCTTTAGCTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db 447 AsnIleLeuArgAlaAlaGluAlaProLeuArgGlnIleValAlaAsnAlaGlyAsp 466
QY 1381 GAAGTTCATTGTTGTAGAAAAAGTTCTGCAACCAAAAGATGTTTGGATTTAATGCT 1440
Db 467 GluProSerValValIleAsnAlaValLysAsnGlyGluGlyAsnPheGlyTyrAsnAla 486
QY 1441 GCATCAGCAGATATGAAGACCTTATTAAAGCTGCTGATGATGATCTTAAAGAGTTACA 1500
Db 487 AlaThrGlyGluTyrGlyAspMetLeuGluMetGlyIleLeuAspProAlaLysValThr 506
QY 1501 CGTATTCATTACAAATGTCAGCATCAGTACGCTCTTACTTCTAATCACTACAGATGCGCT 1560
Db 507 ArgSerAlaLeuGluHisAlaAlaSerValAlaGlyLeuMetLeuThrThrGluCysMet 526
QY 1561 ATTCTCTGAAGAACACAGAACCTTAAAGATATGCTATGCTGCGGTGGTATGGTGGT 1620
Db 527 IleThrAspIleProGluAspLysProAlaAlaPro---AspMetGlyGlyMetGlyGly 545
QY 1621 ATGGGTGGTATG 1632
Db 546 MetGlyGlyMet 549

RESULT 6
US-08-467-822-32
; Sequence 32, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerebaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-822-32
Alignment Scores:
Pred. No.: 5,37e-160 Length: 548
Score: 1741.00 Matches: 353
Percent Similarity: 77.78% Conservative: 74
Best Local Similarity: 64.30% Mismatches: 118
Query Match: 58.66% Indels: 4
DB: 2 Gaps: 2
US-09-077-574A-1 (1-1647) x US-08-467-822-32 (1-548)
QY 1 ATGCTTCTTAAGAAATCCCTTTTGTATGCTAAAGCCCGTGAAGAACTTTCACGAGGTGA 60
Db 1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20
QY 61 GATAAATCTGCAATGCTGTAAAGTAACACTTGGACCTAAAGCCCGTAAATGCTGTTATT 120
Db 21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGTCTTTTGGTTCCTCCAGTTATTACAAAGATGCTGTATCTGTGCAAGAAATTT 180
Db 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60
QY 181 GAATCTGAAGATAAGTTTGAAATATATGGCGCTCAAAATGCTTAAAGAGTAGCTCCCAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTGCTGCTGATGGAACACTACACAGCAACAGCTCTTGCACAGCTATTAT 300
Db 81 AlaAsnAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIle 100
QY 301 CGTGAAGTGTAAACCTTGTAGCAGCTGGTCGTAATCTTATGCGCATTAACCGTGCATA 360
Db 101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLe 120
QY 361 GATAAAGCTGTTGCTGCTGTTACTTAAAGCACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db 121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140
QY 421 CAAAAGAATAGCTCAAGTTGGAAACCTTTCTGCAAACTCTGATACACAAATAGGTAAT 480
Db 141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGAAAGAGGTGTTATTCACAGTTGAGGAGCT 540
Db 161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
QY 541 AAAGGTCTTGAATACATACATAGTAGTGTGTTGAAGGAATGAAGTTGACCTGGCTACTC 600
Db 181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
```


Db 201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220
QY 661 CTTTGTAAAGAGAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTT 720
Db 221 LeuLeuAlaAspLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCCTTATTATGCTGAAGACGTAGAAGTGAAGCACTTGCA 780
Db 241 AlaLysAlaGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTTGTAGTCAATAAGCTCCGTGAGCACTCCAAAGTCTAGCCGTAAAGCTCCCTGGT 840
Db 261 ThrAlaValAsnThrIleArgGlyIleValLysValAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGGTAAAGCTATGCTTGAAGATATGCTATCTCTTACTGAGAGAGCA 900
Db 281 PheGlyAspArgGlyAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300
QY 901 ATATTTGAAGATCTGCTATAAAGCTTGAAGTGTAAAGCTTGTCTCTTTAGCAACAGCT 960
Db 301 IleSerGluGluLeuGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320
QY 961 AAACGTGTAGTATTGACAAAGAAATACTACTATCTGCTGATGCTGCTGAAATCAGAA 1020
Db 321 LysArgValValIleAsnLysAspThrThrIleIleAspGlyValGlyGluGluAla 340
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAAGAACAGCTCAGATTAT 1080
Db 341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360
QY 1081 GATCGTGAAACCTTCAAGACGCTTGTGCAAACTTGTCTGCTGAGTGTATTCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGATCGTGTAGAGATGCTCTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysLysLysAlaArgValGluAspAlaLeu 400
QY 1201 AATGCAACAGACTGGGTGAAGAGGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeu 420
QY 1261 CGCTCCATTAAAGCTCCTTATGATATTAAACCTGCTGATGATGATGATGATGATGAT 1320
Db 421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440
QY 1321 AATPATCATCGCTGCTCTTCTTGAAGAGCCTTTACGTCAAAATTCGTCAAAATTCGTGCTAT 1380
Db 441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460
QY 1381 GAAGGTCTATTGTTAGAAAAGTTCGTGAACCAAGATGCTTTGGAATTAATGCT 1440
Db 461 GluProSerValValAlaAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGGAAATGAAGACCTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 481 AlaThrGluGlyArgMetIleAspMetGlyIleLeuAspProThrLysValThr 500
QY 1501 CGTATTGCAATTAACAAATGACGATCAGTACGCTCTTACTTCTACTACAGAAATGCGCT 1560
Db 501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520
QY 1561 ATTGCTGAAAACAGACACCTTAAAGATATATGCTTATGCTGCTGCTGCTGCTGCTGCT 1614
Db 521 ValThrAspLeu-----ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538
QY 1615 GGTGGTATGGGTATGGCGGTATG 1641
Db 539 GlyGlyMetGlyGlyMetGlyGlyMet 547

RESULT 8

US-08-432-697-32

; Sequence 32, Application US/08432697

; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; POLYPEPTIDES
; TITLE OF INVENTION: 44
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-432-697-32

Alignment Scores:
Pred. No.: 5,37e-160 Length: 548
Score: 1741.00 Matches: 353
Percent Similarity: 77.78% Conservative: 74
Best Local Similarity: 64.30% Mismatches: 118
Query Match: 58.66% Indels: 4
DB: 3 Gaps: 2

US-09-077-574A-1 (1-1647) x US-08-432-697-32 (1-548)
QY 1 ATGGCTCTTAAGAAATCTCTTTTGTATGCTAAAGCCGTGAAATCTTTCCAGAGGTGA 60
Db 1 MetAlaAlaLysValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20
QY 61 GATAAATCTTGAATGCTGTAAAGTAACACTTGGACCTAAAGCCGTAAATGTCGTTATT 120
Db 21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAATGCTTTGGTCCCGAGTATTACAAAGATGGTGTATCTGCTGCAAAAGAAAT 180
Db 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60
QY 181 GAATCTGAAGATAAGTTGAAATATGCGCGCTCAATGGTAAAGAGTAGCTCCCAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATCTGCTGATGCAACTACAAACAGCAACAGCTCTTGCACAGCTATTAT 300
Db 81 AlaAsnAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIle 100

QY 301 CGTGAAGCTGTAACCTTGTAGCAGCTGCTGCTATCTATGCGCATTAACCTGCGCAT 360
Db 101 ThrGluGlyLeuLysAlaValAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAGCTGTTGTTGCTTACTAAAGAACTTAAGCGACATTATCAAGACCTACTCGTGAC 420
Db 121 AspLysAlaValThrAlaAlaValGluGluLysAlaLeuSerValProCysSerAsp 140
QY 421 CAAAGAAATAGCTCAAGTTGGACCACTTTCTGCAAACTCTGTATACAACTAGGTAAT 480
Db 141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160
QY 481 ATCTAGCTGAGCTATCGCTAAAGCTTGAAGAGGAGGTGTATCACAGTTGAGGAAGCT 540
Db 161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
QY 541 AAAGCTCTTGAACCTACATAGATGCTGTTGAAGAAATGAAGTTTGAACGCTGCTACCTC 600
Db 181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCATCTTTGTAACCTTCTGAGAAATGGTTTGTGAACCTTGAACCTTATATC 660
Db 201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220
QY 661 CTTTGTATGAGAAAGATTAAGTACATGAAAGACATGCTACCAATCTTAGAACAGTT 720
Db 221 LeuLeuAlaAspLysIleSerAsnIleArgGluMetLeuProValIleGluAlaVal 240
QY 721 GCTAAGTAAACCGTCCACTCTTATTATTGCTGAAGAGCTGAGAAGTGAAGCACTTGA 780
Db 241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTGTAGCAATAAGCTCGTGGAGCATTCCAAAGTTGTAGCGTAAAGCTCTGCT 840
Db 261 ThrAlaValValAsnThrIleArgGlyIleValLysValAlaAlaValLysAlaProGly 280
QY 841 TTTGCTGAGCGCTGAAGATGCTTGAAGATATTGCTATCTCTTACTGAGGAGCAAGCA 900
Db 281 PheGlyAspArgGlyAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300
QY 901 ATATTGAAGATCGTGGTATAAGCTTGAAGCTTGAAGCTTCTTCTTTAGGAACAGT 960
Db 301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320
QY 961 AAACGTGATATTATGCAAGAAATCTACTATCTGTTGCTGCTGCTGGAATAACGAA 1020
Db 321 LysArgValValIleAsnLysAspThrThrIleIleAspGlyValGlyGluGluAla 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATTAT 1080
Db 341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360
QY 1081 GATCGTGAACAACTTCAAGACCTCTTGAACAACTTGTGTGAGTAGCTGCTTATCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyValAlaValIleLys 380
QY 1141 GTTGAGCTGCTACTGAACCTGAATGAAGAGAGAGGATCGTGTAGAACATGCTCTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
QY 1201 AATGCAACAGAGCTGCGTTGAAGAGGTATTGTCCTGCTGCTGCTGCTCTTTGTC 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuIle 420
QY 1261 CGCTCCATTAAGCTCTGATGATATTAACCTGCTGATGATGATCAACCTGCTGACCT 1320
Db 421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440
QY 1321 AATATCATCCGCTGCTCTTGAAGAGCTTTACGTCAATTTGCTGCAATCTGCTGCTAT 1380
Db 441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

QY 1381 GAAGTCTCTATTGTTAGAAAAAGTTCTGTGAACCAAAAGATGTTTGGATTATATGCT 1440
Db 461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGAGAAATATGAAGACCTTATTAAAGCTGTGTGTCATGATCTTAAAAAGTTACA 1500
Db 481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500
QY 1501 CGTATTGCAATACAAAATGCAGCATCACTAGCTCTCTTACTTCTAACTACAGAAATGCGCT 1560
Db 501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520
QY 1561 ATTGCTGAAAAACAGAACCTTAAAAAGATATGCTATGCTCTGCGTGGC-----GGTGGTATG 1614
Db 521 ValThrAspLeu-----ProLysAsnAspAlaAlaAspLeuGlyAlaGlyGlyMet 538
QY 1615 GGTGTATGGTGGTATGACGGTATG 1641
Db 539 GlyGlyMetGlyGlyMetGlyGlyMet 547

RESULT 9
US-08-466-248-32
; Sequence 32, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-466-248-32

Alignment Scores:

Pred. No.: 5,37e-160 Length: 548
Score: 1741.00 Matches: 353
Percent Similarity: 77.78% Conservative: 74
Best Local Similarity: 64.30% Mismatches: 118
Query Match: 58.66% Indels: 4
DB: 3 Gaps: 2

US-09-077-574A-1 (1-1647) x US-08-466-248-32 (1-548)

QY 1 ATGCTCTTAAAGAAATCTTTTTCATGCTAAGCCCGTGAAGAACTTTCACGAGGTGA 60
DB 1 MetAlaAlaLysAspValPheGlyAsnMetGlyValLysMetLeuArgGlyVal 20
QY 61 GATAAACCTTCAATGCTGTTAAAGTAACTTGGACCTTAAGCCGCTAATGCTGTTATT 120
DB 21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAAGTCTTTTGGTCCCGAGTTATTACAAAAGATGCTGTATCTGTTCGCAAGAAAT 180
DB 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluLe 60
QY 181 GAACCTGAAGTAAAGTTGAAATATATGGCGCTCAAAATGCTTAAAGAGTAGCTCCCAAA 240
DB 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATGCTGTGTGATGAGTACACACACACAGTCCCTTGGACAGCTATTAT 300
DB 81 AlaAsnAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIle 100
QY 301 CGTGAAGCTGTAACCTTTGAGCAGCTGCTGTAATCTTATGGCCATTAAACGTCGATA 360
DB 101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyLe 120
QY 361 GATAAGCTGTTGCTGTTGCTTACTTAAAGAACTTAAGCGACATTACAAAGCCTACTCGTGAC 420
DB 121 AspLysAlaValThrAlaAlaValGluLeuLysAlaLeuSerValProCysSerAsp 140
QY 421 CAAAAGAAATAGCTCAAGTTGGNACCATTTCTGCAAACTCTGATACAACTAGGTAAT 480
DB 141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160
QY 481 ATCATAGCTCAAGCTATGCTGCTTAAAGTTGAAAGAGGCTGTATCACAGTTGAGGAAGCT 540
DB 161 LeuIleAlaGluAlaMetAspLysValGlyLysGlnGlyValIleThrValGluAspGly 180
QY 541 AAAGGCTTGAAGAACTACATTAGATGTTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
DB 181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATCTTTGTAATACTCTCTGAGAAATGTTGTGAACCTTGATTAACCTTATATC 660
DB 201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220
QY 661 CTTTGTAATGAGAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGACAAAGTT 720
DB 221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
DB 241 AlaLysAlaGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTTGTTAGTCAATAGCTCCGTGGAGCACTCCAAGTTGACCGGTAAAGCTCTCGGT 840
DB 261 ThrAlaValValAsnThrIleArgGlyIleValLysValAlaValLysAlaProGly 280
QY 841 TTTTGGTGAAGCGGTAAAGCTAGCTGAGATATTGCTATCTCTACTGAGGAGAGCA 900
DB 281 PheGlyAspArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyThrVal 300
QY 901 ATATTGAGATCGTGGTATAAAGCTTGAAGCTTGAAGCTTGTCTTCTTTAGGAACAGCT 960

DB 301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320
QY 961 AAACGTGTAGTTATGCAAAAGAAATACTACTATCGTTGATGCTGCTGGAATACAGAA 1020
DB 321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTSCACAAATTCGAAGAAACAAAGCTCAGATTAT 1080
DB 341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluAlaThrSerAspTyr 360
QY 1081 GATCCTCAAAAACCTTCAAGAACGCTTTCGAAACCTTGTTCGTGAGTAGCTGTTATCCAT 1140
DB 361 AspArgGluLysLeuGlnArgValAlaLysLeuAlaGlyValAlaValIleLys 380
QY 1141 GTTCGAGCTGCTACTGAAACCTGAATGAAAGAGAGAGGATCGCTAGAGAGATGCTCTA 1200
DB 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
QY 1201 AATGCAACAAGCTCGGTTGAGAGAGGTATTCCTCGTGTGCTGCTGCTTTGTC 1260
DB 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuIle 420
QY 1261 CGTCCATTAAAGTCTTGTATGATATTAAACCTGCTGATGATGATGAACCTTGTGGACTT 1320
DB 421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyLe 440
QY 1321 AATATCATCGTCTGCTTCTTGTGAAGCCTTTAGTCGTCGTCGTCGTCGTCGTCAT 1380
DB 441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460
QY 1381 GAAGGTTCTATTGTTGAGAAAGTTCTGTGACCAAAAGATGCTTTGATTAATGCT 1440
DB 461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGCAATATGAAGACCTTATTAAAGCTGCTGATGATGATCTCTAAAGAGTTACA 1500
DB 481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500
QY 1501 CGTATTGATTCAAAATGCAAGCATCAGTAGCCCTCTTACTTCTAAGTACAGAAATGCGCT 1560
DB 501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520
QY 1561 ATTGCTCAAAAACCAACCTTAAAGATATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 1614
DB 521 ValThrAspLeu-----ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538
QY 1615 GGTGGTATGGTGGTATGAGCGGTATG 1641
DB 539 GlyGlyMetGlyGlyMetGlyGlyMet 547

RESULT 10

US-08-470-260-6

Sequence 6, Application US/08470260

Patent No. 6077706

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello

APPLICANT: Bugnoli, Massimo

APPLICANT: Telford, John

APPLICANT: Macchia, Giovanni

APPLICANT: Rappuoli, Rino

TITLE OF INVENTION: Helicobacter Pylori Proteins Useful

TITLE OF INVENTION: for Vaccines and Diagnostics

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESS: Chiron Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: California

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0316.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-260-6

Alignment Scores:
Pred. No.: 2,65e-158 Length: 546
Score: 1723.50 Matches: 350
Percent Similarity: 77.4% Conservative: 73
Best Local Similarity: 64.10% Mismatches: 120
Query Match: 58.0% Indels: 3
DB: Gaps: 2

US-09-077-574A-1 (1-1647) x US-08-470-260-6 (1-546)

QY 7 TCTAAGAAATCTTTTTCATGCTAAAGCCGCTGAAAACTTTCACAGGTGTAGATAA 66
DB 2 AlaLysGluLeuLysPheSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21
QY 67 CTTGCAAAATGCTTTAAAGTAAACACTTGACCTTAAAGCGCGTAATTCGTTTAAAG 126
DB 22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuLeuGlnLys 41
QY 127 TCTTTTGGTTCCTCCAGTATTACAAAAGATGGTGTATCTGTGCAAAAGAAATGAAC 186
DB 42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluLeuLeu 61
QY 187 GAAGATAAGTTTGAAATATGGCGCTCAATGGTTAAAGAACTAGCTCCCAAACTAGC 246
DB 62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81
QY 247 GATATTGCTGGTGGATGCACTACACAGCAACAGTCTTGCACAAAGCTATTATCGTGA 306
DB 82 AspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaLysSerIlePheLysGlu 101
QY 307 GGTTGTAAGTCTAGCAGCTGTGTAAATCCCTATGGCCATTAAAGCTGCATAGATAA 366
DB 102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121
QY 367 GCTGTTGTTGCTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAA 426
DB 122 AlaAlaGluAlaIleLeuAsnGluLeuLysLysAlaSerLysValGlyGlyLysGlu 141
QY 427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACTATGATATATCAT 486
DB 142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeu 161
QY 487 GCTGAAGCTATGGCTAAAGTTGAAAGAGGTGTATATCATAGTTGAGGAGCTAAAGGT 546
DB 162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181
QY 547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGCTCTCTCTCCA 606

DB 182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201
QY 607 TACTTTGTAACCTAATCTGAGAAAATGCTTTGTGAACCTTGATAACCCCTATATCTTTGT 666
DB 202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221
QY 667 AATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAGTTCTCTAAA 726
DB 222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys 241
QY 727 GTAAACCGTCCACTCTTATTATTCTGAAGACGTAGAGGTGAAGCTTCCACACTT 786
DB 242 GluGlyLysProLeuLeuIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261
QY 787 GTAGTCAATAAGCTCGGTGGACCTCCCAAGTTGTAGCGTAAAGAGCTCTCTGTTTGT 846
DB 262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaValLysAlaProGlyPheGly 281
QY 847 GAACCGCGTAAAGCTATGCTTGAAGATATTGCTATCTTACTTGGAGGAGAACATATT 906
DB 282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301
QY 907 GAAGATCGTGTATAAAGCTTGAAATGTAACTGTCTTCTTTAGGACACAGCTAAACGT 966
DB 302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321
QY 967 GTAGTATTGCAAGAAATACTACTATCTGTTGATGTTGCTGGAATAATCAGAAATATT 1026
DB 322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341
QY 1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGCT 1086
DB 342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361
QY 1087 GAAAACTTCAAGAACGCTTTCGCAAACTGTTGGTGGAGTAGCTGTTTATCCATGTTGA 1146
DB 362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381
QY 1147 GCTGCTACTGAACTGAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
DB 382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401
QY 1207 ACAAGAGCTCGCTTGAAGAGAGTATTGTCCTCGTGGTGGTGGTGGTGGTGGTGGT 1266
DB 402 ThrLysAlaAlaValGluGluGluGlyIleValIleGlyGlyAlaAlaLeuIleArgAla 421
QY 1267 ATTAAGTCTCTGATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGAT 1326
DB 422 AlaGlnLysValHis-----LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439
QY 1327 ATCCGTCGTTCTTGAAGAGCGCTTACGTCAAAATGCTGCAATGCTGCAATGCTGCAAT 1386
DB 440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459
QY 1387 TCTATTGTTGTACAAAAGTTTCGTGAACCAAAAGATGTTTGGATTAAATGCTGCATCA 1446
DB 460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479
QY 1447 GGAATATGAAACCTTATTAAAGCTGGTGTGCTTATGCTTCTTAAAGTACACGTTAT 1506
DB 480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499
QY 1507 GCATTCAAAATCCAGCATCAGTAGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1566
DB 500 AlaLeuGlnAsnAlaValSerValSerLeuLeuLeuThrThrGluAlaThrValHis 519
QY 1567 GAAAAACCAACCTAAAAAAGATATGCTCTATGCTGGC---GCTGATGGGTGGTATG 1623
DB 520 GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet 539
QY 1624 GGTGGTATGACCGGTATG 1641

RESULT 12

US-08-466-662-6
; Sequence 6, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappucci, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-466-662-6

Alignment Scores:
Pred. No.: 2,65e-158 Length: 546
Score: 1723.50 Matches: 350
Percent Similarity: 77.47% Conservative: 73
Best Local Similarity: 64.10% Mismatches: 120
Query Match: 58.07% Indels: 3
DB: 2 Gaps: 2

US-09-077-574A-1 (1-1647) x US-08-466-662-6 (1-546)

QY	7	TCTAAGAAATCCTTTTTCATGCTAAAGCCGCTGAAAAAATTTCACAGGTGAGATAAA	66
DB	2	AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln	21
QY	67	CTTGCAATGCTGTTAAAGTAACACTTGACCTAAAGGCGGTAATCTGCTTATTGAAAG	126
DB	22	LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys	41
QY	127	TCTTTTGGTCCCGAGTTATCAAAAGATGGTGTATCTGCTCAAAAGAAATTTGAATTT	186
DB	42	SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu	61
QY	187	GAAGATAAGTTTGAAATATATGGCGCTCAAAATGTTTAAAGAGTAGCTCCCAAACTAGC	246
DB	62	SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla	81
QY	247	GATATGCTGCTGATGGAACCTACACAGCAAGTCTTGCACAGCTATTATTCGTGAA	306
DB	82	AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu	101
QY	307	GGTGTAATACTCTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	366
DB	102	GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys	121
QY	367	GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	426
DB	122	AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysValGlyGlyLysGlu	141
QY	427	GAATAGCTCAAGTTGGAACCATTTCTGCAACTCTGATACACAACTAGTAATATCATA	486
DB	142	GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle	161
QY	487	GCTGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	546
DB	162	AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluAlaLysGly	181
QY	547	CTTGAAACTACATTAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	606
DB	182	IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro	201

QY	607	TACTTTGTAACATACTCTGAGAAAAATGGTTTGTGAACCTTGATAACCTTATATCTCTTGT	666
DB	202	TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu	221
QY	667	AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTTGCTAAA	726
DB	222	ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys	241
QY	727	GTAACCCGCTCCACTCTTATTATTCTGTAAGACGTAGAGGTGAAGCACTTCACACATT	786
DB	242	GluGlyLysProLeuLeuIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu	261
QY	787	GTAGTCAATAAGCTCCGTGGAGCACTCCAAGGTGTAGCGTAAAGCTCCTCGTTTGT	846
DB	262	ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly	281
QY	847	GAACCCGCTAAAGCTATGCTGAAGATATGCTATCTCTTACTTGGAGGAGCAATATTT	906
DB	282	AspArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer	301
QY	907	GAAGATCGTGTATTAAGCTTCAAAATGTAAGCTTCTCTTTCAGCAACAGCAACCT	966
DB	302	GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg	321
QY	967	GTAGTATTGACAAAGAAATACTACTATCTGTTGATGCTGCTGGAANAATCAGAAATAT	1026
DB	322	IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspVal	341
QY	1027	AAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATTATGCT	1086
DB	342	LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys	361
QY	1087	GAAGAACTTCAAGACGCTCTTCAAACTTGTGTTGGAGTAGCTGTATCCATGTTGGA	1146
DB	362	GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly	381
QY	1147	GCTGCTACTGAAACTGAAATGAAAGAGAGATCGTGTAGAGATGCTCTTAAATGCA	1206
DB	382	AlaAlaSerGluValGluMetLysGluLysAspArgValAspAlaLeuSerAla	401
QY	1207	ACAAGAGCTGCGGTTGAAAGAGTATGTCCTCCCTGGTGGTGGTACTGCTTTTCCGCTCC	1266
DB	402	ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla	421
QY	1267	ATTAAAGTCTTCATCATATTAAACCTGCTGATGATGATGATGCTGCTGCTGCTGCTGCT	1326
DB	422	AlaGlnLysValHis-----LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle	439
QY	1327	ATCCGCTGCTTCTTGAAGAGCCTTTTACGCTCAAAATTCGCTGCAAAATCTGCTATGAGGT	1386
DB	440	IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly	459
QY	1387	TCTATTGTTGTAAGAAAGTTCGTGAACCAAGAGTGGTTTGGATTATGCTGATCA	1446
DB	460	GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn	479
QY	1447	GGAGATATCAAGACCTTATTAAAGCTGGTGTCTATTCATCTCTTAAAGTTACACGTTAT	1506
DB	480	GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle	499
QY	1507	GCATTACAAATACAGCATCAGTAGCTCCTTACTTCTTACTTCTTACTTCTTACTTCTTACT	1566
DB	500	AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis	519
QY	1567	GAAGAAACCAAGCACTTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1623
DB	520	GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet	539
QY	1624	GCTGGTATGACCGGTATG 1641	
DB	540	GlyGlyMetGlyGlyMet 545	

RESULT 13

US-08-256-847C-1
; Sequence 1, Application US/08256847C
; Patent No. 6403099
; GENERAL INFORMATION:
; APPLICANT: Rappuoli, Rino
; APPLICANT: Costantino, Paolo
; APPLICANT: No. 6403099elli, Francesco
; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
; FILE REFERENCE: CHIR-0042
; CURRENT APPLICATION NUMBER: US/08/256.847C
; CURRENT FILING DATE: 1994-11-01
; PRIOR APPLICATION NUMBER: PCT/EP93/00516
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: F192A000058
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 546
; TYPE: PRT
; ORGANISM: H. pylori
US-08-256-847C-1

Alignment Scores:

Pred. No.:	2,65e-158	Length:	546
Score:	1723.50	Matches:	350
Percent Similarity:	77.47%	Conservative:	73
Best Local Similarity:	64.10%	Mismatches:	120
Query Match:	58.07%	Indels:	3
DB:	4	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-256-847C-1 (1-546)

QY	7	TCTAAGAAATCCTTTTGATGCTAAAGCCGCGTGAAGAAATCTTCCAGAGGTGATAGATAA	66
DB	2	AlalysGluilleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln	21
QY	67	CTTGCAATGCTGTAAAGTAACACTTGGACCTAAGCCGTATGCTATGCTTAAAGAAATGAA	126
DB	22	LeuHisaspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuileGlnLys	41
QY	127	TCCTTTGCTCCCGAGTATTACAAAAGATGCTGTATCTGTGCAAAAGAAATGAACTT	186
DB	42	SerTyrGlyAlaProSerIleThrLysaspGlyValSerValAlaLysGluileGluLeu	61
QY	187	GAAGATAAGTTTGAATATGGCGCTCAATAGTTTAAAGAGTAGTCCCAAAACTAGC	246
DB	62	SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaLaserLysThrAla	81
QY	247	GATATTGCTGCTGATGAACACTACACAGCAACAGTCTTGCACAACTATTATTCGTGAA	306
DB	82	AspAlaAlaGlyaspGlyThrThrAlaThrValLeuAlaLysSerIlePheLysGlu	101
QY	307	GGTGTAACCTCTAGCAGCTGGTGTATCTTATCCCTATGGCCATTAACCGTGCATAGATAA	366
DB	102	GlyLeuArgAsnIleThrAlaGlyAlaAsnProileGluValLysArgGlyMetAspLys	121
QY	367	GCTGTGCTGCTGTTACTAAGAACTAAGCGACATTAACAAGCCTACTCGTGACCAAAA	426
DB	122	AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu	141
QY	427	GAAATAGCTCAAGTGAACATTTCTGCAAACTCTGATACCAACAAATAGTAATATCATA	486
DB	142	GluileThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuile	161
QY	487	GCTGAAGCTATGGCTAAAGTTGAAAAGGAGGTCTTATCAAGTTGAGGAAGCTTAAAGGT	546
DB	162	AlaaspAlaMetGluLysValGlyLysaspGlyValIleThrValGluGluAlaLysGly	181
QY	547	CTTGAAACTACATTAGATGGTTGAAAGGAATGAAGTTTGACCGTGGCTCTCTCCA	606
DB	182	IleGluaspGluLeuaspValValGluGlyMetGlnPheaspArgGlyTyrLeuSerPro	201

QY	607	TACTTTCTAATCTCTGAAAATGGTTTGTGAACTTGATAACCTTATATCTTTGCT	666
DB	202	TyrPheValThrAsnAlaGluLysMetThrAlaLeuAspAsnAlaLafyrIleLeuLeu	221
QY	667	AATGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAAGTTGCTAAA	726
DB	222	ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys	241
QY	727	GTAACCGTCCACTCTTATTATTGCTGAAGACCTAGAGGTGAAGCACTTGCAACACT	786
DB	242	GluGlyLysProLeuLeuIleAlaGluaspIleGluGlyGluAlaLeuThrThrLeu	261
QY	787	GTAGTCAATAGCTCCGCTGGAGCACTCCAAGTTCTAGCCGTAAAGCTCTGGTTTGGT	846
DB	262	ValValAsnLysLeuArgGlyValLeuAsnIleAlaValLysAlaLafyrPheGly	281
QY	847	GAACGCGTAAAGCTATGCTTGAAGATATTGCTACTCTGAGGAGAGCAATATTT	906
DB	282	AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer	301
QY	907	GAAGATCGTGTATAAAGCTTGAAGATGTAAAGCTTGTCTTTTGGAGCAAGCTTAAAGT	966
DB	302	GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg	321
QY	967	GTAGTTATTGACAAAGAAATACTACTATCGTTGATGCTGCTGGAATCAGAGATATT	1026
DB	322	IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspVal	341
QY	1027	AAAGCTCGAGTTAAACAAATTCGTGCACAAATTCAGAAACAAAGCTCAGATTATGATCGT	1086
DB	342	LysAspArgValAlaGlnIleLysThrGluIleAlaSerThrThrSerAspLysAspLys	361
QY	1087	GAAAACTTCAAGAACCTCTTGTGCAAACTTGTGGTGAGTAGCTGTATCCATGTTGGA	1146
DB	362	GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly	381
QY	1147	GCTGCTACTCAAACTGAATGAAAGCAAGCAAGATCGTGTAGAGAGCTCTTAAATGCA	1206
DB	382	AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAlaLeuSerAla	401
QY	1207	ACAAGACTCGGTTGAAGAGGTATTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC	1266
DB	402	ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyAlaAlaLeuIleArgAla	421
QY	1267	ATTAACTCTCTGATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGATC	1326
DB	422	AlaGlnLysValHis-----LeuAsnLeuHisAspAspGluLysValGlyTyrGluile	439
QY	1327	ATCCGCTGCTCTCTGAGAGAGCTTTTACGTCAAATGCTGCAATGCTGCAATGCTGCAAGGT	1386
DB	440	IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly	459
QY	1387	TCATTCTTCTGAGAAAAGTTCTGTAACCAAAAGATGGTTTGTGATTTAATGCTGATCA	1446
DB	460	GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn	479
QY	1447	GGAGATATGAAGACCTTATTAAAGCTGGTGTCTATTGATCCTTAAAGTTTACACGTTAT	1506
DB	480	GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle	499
QY	1507	GCATTCAAAATCGAGCATCAGTACCTCTTACTTCTTAATCAGATCGGCTATTGCT	1566
DB	500	AlaLeuGlnAsnAlaValSerValSerSerLeuLeuThrThrGluAlaThrValHis	519
QY	1567	GAATAACCAACCTAAAAAGATATGCTCTATGCTGGC---GTCGTATGGGTGGTATG	1623
DB	520	GluileLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet	539
QY	1624	GGTGGTATGGACGGTATG	1641
DB	540	GlyGlyMetGlyGlyMet	545

RESULT 14

US-08-256-847C-7
 ; Sequence 7, Application US/08256847C
 ; Patent No. 6403099
 ; GENERAL INFORMATION:
 ; APPLICANT: Rappuoli, Rino
 ; APPLICANT: Costantino, Paolo
 ; APPLICANT: No. 6403099elli, Francesco
 ; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
 ; FILE REFERENCE: CHIR-0042
 ; CURRENT APPLICATION NUMBER: US/08/256,847C
 ; CURRENT FILING DATE: 1994-11-01
 ; PRIOR APPLICATION NUMBER: PCT/EP93/00516
 ; PRIOR FILING DATE: 1993-03-08
 ; PRIOR APPLICATION NUMBER: FI92A000058
 ; PRIOR FILING DATE: 1992-03-06
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 7
 ; LENGTH: 546
 ; TYPE: PRT
 ; ORGANISM: H. pylori
 US-08-256-847C-7

Alignment Scores:
 Pred. No.: 2,656-158 Length: 546
 Score: 1723.50 Matches: 350
 Percent Similarity: 77.47% Conservative: 73
 Best Local Similarity: 64.10% Mismatches: 120
 Query Match: 58.07% Indels: 3
 DB: 4 Gaps: 2

US-09-077-574A-1 (1-1647) x US-08-256-847C-7 (1-546)

QY	7	TCTAAAGAAATCCTTTTTCATGCTAAAGCCCGTGAAGAAATCTTCACAGGTGATGATAA	66
DB	2	AlaLysGluLeuLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln	21
QY	67	CTTGCAAACTGCTGTTAAAGTAACACTTGACCTTAAAGGCGTAATGCTGTTATTTGAAAG	126
DB	22	LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuLeuGlnLys	41
QY	127	TCCTTTGGTTCCTCCAGCTATTACAAAAGATGGTGATCTGTTCCAAAAGAAATGGAACIT	186
DB	42	SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluLeuGln	61
QY	187	GAAGATAAGTTTGAAATATGGCGCTCAAAATGGTTAAAGAGTAGTCCCAAACTAGC	246
DB	62	SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla	81
QY	247	GATATGCTGGTGAAGTCAACAGCAAGTCTTCACAGCTATTATTCGTGAA	306
DB	82	AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu	101
QY	307	GGTGTAACCTCTAGCAGCTGCTGTAATCTTATGGCCATTAAACCTGGCATAGATAA	366
DB	102	GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys	121
QY	367	GCTGTTGCTGCTTACTAAAGAACTAAGCAGCATTAACAAAGCCTACTCGTACCAAAA	426
DB	122	AlaAlaGluAlaIleLeuAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu	141
QY	427	GAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACACAAATAGTAATATCATA	486
DB	142	GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle	161
QY	487	GCTGAAGCTATGGCTAAAGTTGAAAAGAGGTGTTATCACAGTTGAGGAGCTTAAAGGT	546
DB	162	AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly	181
QY	547	CTTGAAACTACATTAGATGGTTGAGGAATGAAGTTACCGTGCCTACCTCTCTCCA	606
DB	182	IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro	201

QY	607	TACTTTGTAACTAATCCTCAGAAAAATGGTTTGTGAACCTTGATAAACCTTATATCTCTTGT	666
DB	202	TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu	221
QY	667	AATGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACACAGTTGCTAAA	726
DB	222	ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys	241
QY	727	GTAACCCGCTCACTCTTATTTCTGAAGCTAGAGGTGAGAGCTTGCACACACTT	786
DB	242	GluGlyLysProLeuLeuIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu	261
QY	787	GTAGTCATAAAGCTCCGTGGAGCCTCCCAAGTTGTAGCCGTAAAAAGCTCTCTGTTTGT	846
DB	262	ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly	281
QY	847	GAACCCGCTAAAGCTATGCTTGAAGATATTGCTATCTCTTACTGGAGGAGCAATATTT	906
DB	282	AspArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer	301
QY	907	GAAGATCGTGTATAAAGCTTGAAGATGTAAAGCTTCTCTTTAGGAACACTAAACGT	966
DB	302	GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg	321
QY	967	GTAGTTATTGACAAAAGAAATACTACTATCTGTTGATGGTGGGAAATCAGAAATATT	1026
DB	322	IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal	341
QY	1027	AAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGAT	1086
DB	342	LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys	361
QY	1087	GAAGAACTTCAAGACGCTTTCACAACTTGTGTGGAGTAGCTCTTATCCATGTTGGA	1146
DB	362	GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly	381
QY	1147	GCTGCTACTCAAACTGAAATGAAAGAGAGATCGTGTAGAGATGCTCTTAAATGCA	1206
DB	382	AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla	401
QY	1207	ACAAGAGCTCGGTTGAAGAGTATTGTCCTCGTGGTGGTACTGCTGTTTCCGCTCC	1266
DB	402	ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla	421
QY	1267	ATTAAAGTCTTCATCATATTAAACCTGCTGATGATGATGATGCTGCTGACCTTAATATC	1326
DB	422	AlaGlnLysValHis-----LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle	439
QY	1327	ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCCTCAAAATTCCTGCAAAATGCTGCTATGAGGT	1386
DB	440	IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly	459
QY	1387	TCTATTGTTGTAAGAAAGTTCGTGACCAAAAGATGGTTTGGATTATTCGTGATCA	1446
DB	460	GlyValValValAsnGluValGluLysHisGlyGlyHisPheGlyPheAsnAlaSerAsn	479
QY	1447	GGAGATATCAAGACCTTATTAAAGCTGGTGCATTCATTCATCTTAAAGATTTACACGATT	1506
DB	480	GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle	499
QY	1507	GCATTACAAAATCAGCATCAGTAGGCTCTTACTTCTTAACTACAGAATCGCTATTGCT	1566
DB	500	AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis	519
QY	1567	GAAAAACCAAGCACTTAAAAAGATATGCCTATCGCTGCC---GGTGTATCGGTGGTATG	1623
DB	520	GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet	539
QY	1624	GGTGGTATGACCGGTATG 1641	
DB	540	GlyGlyMetGlyGlyMet 545	

RESULT 15

US-08-467-822-33
Sequence 33, Application US/08467822
Patent No. 5843460

GENERAL INFORMATION:

APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,822

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/447,177

FILING DATE: 19-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/432,697

FILING DATE: 02-MAY-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03495, 0137-02000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 544 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-467-822-33

Alignment Scores:

Pred. No.:	4,62e-158	Length:	544
Score:	1721.00	Matches:	351
Percent Similarity:	78.23%	Conservative:	73
Best Local Similarity:	64.76%	Mismatches:	114
Query Match:	57.99%	Indels:	4
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-467-822-33 (1-544)

Qy	1	ATGCTTCTAAAGAAATCCTTTTGTATGCTAAAGCCGCTGAAACCTTTCACGAGGTGA	60
Db	1	MetAlaAlaLysAsnIleLysTyrAsnGluAspAlaArgLysLysIleHisLysGlyVal	20
Qy	61	GATAAAGCTTCAATGCTGTTAAAGTAACTTGGACCTTGGACCTAAGCCGCTAATGCTGTTAT	120
Db	21	LysThrLeuAlaGluAlaValLysValThrLeuGlyProLysGlyArgHisValValle	40

Qy	121	GAAGAAGCTCTTTGGTTCCCAAGTTATTACAAAGATGGTGTATCTGTGTCGAAGAAT	180
Db	41	AspLysSerPheGlySerProGlnValThrLysAspGlyValThrValAlaLysGluIle	60
Qy	181	GAAGTTGAAGATAAGTTTGAATAATATGGCGCTCAAATGGTTTAAAGAACTAGCTCCCAA	240
Db	61	GluLeuGluAspLysHisGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys	80
Qy	241	ACTAGCGATATTGCTGGTGTAGTGAAGAACTACACAGCAACAGTCCTTGCAACAGCTATT	300
Db	81	ThrAlaAspLysAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGluAlaIleTyr	100
Qy	301	CGTGAAGCTGTAAAGAACTTGTAGCAGCTGCTGTAATCTCTAGCCATTAAGCTGGCATA	360
Db	101	SerGluGlyLeuArgAsnValThrAlaGlyAlaAsnProMetLeuAspLysArgGlyIle	120
Qy	361	GATAAAGCTGTTGTTGCTGTTTACTAAAGAACTAGCGACATTAACAAAGCCTACTGTCAC	420
Db	121	AspLysAlaValLysValValValAspGluIleLysLysIleSerLysProValGlnHis	140
Qy	421	CAAAAAGAAATAGCTCAAGTTGGAACTTCTGCAAACTCTGTATACAACTAGCTTAAT	480
Db	141	HisLysGluIleAlaGlnValAlaThrIleSerAlaAsnAspAlaGluIleGlyAsn	160
Qy	481	ATCATAGCTGAAGCTATGCTAAAGTTGGAAGAGGTGTATCACAGTTTCAGGAAGCT	540
Db	161	LeuIleAlaGluAlaMetGluLysValGlyLysAsnGlySerIleThrValGluGluAla	180
Qy	541	AAAGTCTTGAAGAACTACATTAGATGTGTTGAAGGAATGAAGTTTACCGTGGCTACCTC	600
Db	181	LysGlyPheGluThrValLeuAspValValGluGlyMetAsnPheAsnArgGlyTyrLeu	200
Qy	601	TCTCATACCTTTGTAATCTTCCCTGAGAAATGGTTTGTGAACCTGTGATAACCTTATATC	660
Db	201	SerSerTyrPheSerThrAsnProGluThrGlnGluCysValLeuGluGluAlaLeuVal	220
Qy	661	CTTTGTAATGAGAAAGATTTACTAGCATGAAGACATGCTACCAATCTTGAACACAGTT	720
Db	221	LeuIleTyrAspLysLysIleSerGlyIleLysAspPheLeuProValLeuGlnGlnVal	240
Qy	721	GCTAAAGTAAACCGTCCACTCTTTATTATTGCTGAAGACGTAGAGGTGAAGCACTTGC	780
Db	241	AlaGluSerGlyArgProLeuLeuIleAlaGluAspIleGluGlyGluAlaLeuAla	260
Qy	781	ACATTGTAGTCAATAAGCTCCGCTGGAGACTCCAGTTGTAGCCGTAAAGCTCTGGT	840
Db	261	ThrLeuValValAsnArgLeuArgAlaGlyPheArgValCysAlaValLysAlaProGly	280
Qy	841	TTTGTGTAACGCGCTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGAGGAGAGCA	900
Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnLeu	300
Qy	901	ATATTTGAAGATCGTGTATATAAGCTTTGAAAATCTTAAGCTTGTCTTCTTGAAGCAAGCT	960
Db	301	IleSerGluGluLeuGlyMetLysLeuGluAsnThrThrLeuAlaMetLeuGlyLysAla	320
Qy	961	AAAGCTGTAGTTATTGACAAAGAAATATCTATTCCTTGTATGCTGTGCTGGAATATGAA	1020
Db	321	LysLysValIleValSerLysGluAspThrThrIleValGluGlyLeuGlySerLysGlu	340
Qy	1021	GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGTCACAAATTCGAAGAAACA	1080
Db	341	AspIleGluSerArgCysGluSerIleLysLysGlnIleGluAspSerThrSerAspTyr	360
Qy	1081	GATCGTGAAGAACTTCAAGAACTGCTTGTGCAAACTTCTGCTGGAGTAGCTGTTATCCAT	1140
Db	361	AspLysGlnLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleArg	380
Qy	1141	GTTGGAGCTCTACTGAACTGAAATGAAAGAGAGGATCGTGTAGAGATGCTCTA	1200
Db	381	ValGlyAlaAlaThrGluIleGluMetLysGluLysLysAspArgValAspAspAlaGln	400
Qy	1201	AATGCAACAAGAGCTCGCGTTGAAGAGGTATTGTCTCCTCGTGTGCTACTGCTTTTGT	1260

[illegible]

Search completed: January 28, 2004, 13:30:55
Job time : 46.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 28, 2004, 13:17:56 ; Search time 52 Seconds
(without alignments)
10054.718 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaagaatactct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107963 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=cgna2.1/USPTO.spool_p/US09077574/runat 28012004.125845.11596/app_query.fasta_1.1799
-DB=A_Geneseq 19Jun03 -GFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptn -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09077574 @CGN 1.1.81 @runat 28012004.125845.11596 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGHOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneexp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2708	91.2	548	18	AAW16678
2	1864.5	62.8	545	20	AAV23930
3	1848	62.3	549	20	AAV23909
4	1825.5	61.5	547	22	AAE69060
5	1820.5	61.3	547	16	AAE67381
6	1819	61.3	574	22	AAE69061
7	1814.5	61.1	545	20	AAV23915
8	1812.5	61.1	544	21	AAV75747
9	1805.5	60.8	544	24	ABP77245
10	1769.5	59.6	544	23	AAE22070
11	1750.5	59.0	544	21	AAV75745
12	1749	58.9	545	23	ABB99014
13	1746	58.8	545	21	AAE19080
14	1746	58.8	548	19	AAE65070
15	1746	58.8	548	20	AAE95001
16	1746	58.8	548	22	AAE50536
17	1746	58.8	548	23	ABB99010
18	1744	58.8	545	23	ABB99013
19	1743.5	58.7	544	20	AAV23905
20	1742	58.7	545	23	ABB99009
21	1742	58.7	545	23	ABB99012
22	1742	58.7	545	23	ABB99016
23	1741	58.7	548	19	AAE61378
24	1741	58.7	548	21	AAV99621
25	1739	58.6	545	23	ABB99015
26	1738	58.6	545	23	ABB99011
27	1736.5	58.5	548	20	AAV23908
28	1735	58.5	548	16	AAE67382
29	1734	58.4	545	23	ABB99017
30	1733	58.4	545	23	AAV23901
31	1732	58.4	542	23	ABB49241
32	1729.5	58.3	544	23	AAV23903
33	1723.5	58.1	546	14	AAE41661
34	1723.5	58.1	546	14	AAE41200
35	1723	58.1	544	12	AAE13337
36	1720.5	58.0	546	19	AAE98776
37	1720	58.0	548	23	ABB99019
38	1720	58.0	548	23	ABB99021
39	1719	57.9	548	23	ABB99018
40	1718	57.9	544	16	AAE67383
41	1717	57.9	548	23	ABB99020
42	1716.5	57.8	547	23	AAU76192
43	1716.5	57.8	547	24	ABG74586
44	1710	57.6	544	22	AAE11757
45	1710	57.6	544	23	ABB94272

ALIGNMENTS

RESULT 1
AAW16678
ID AAW16678 standard; Protein; 548 AA.
XX
AC AAW16678;
XX
XX
DT 20-AUG-1997 (first entry)
XX
DE Lawsonia intracellularis GroEL.
XX
KW Intestinal disease; porcine proliferative enteropathy; vaccine;
KW GroEL; immunotherapy; antibody; diagnosis; heat shock protein.
XX
OS Lawsonia intracellularis.
XX
FN WO9720050-A1.
XX
PD 05-JUN-1997.

XX 29-NOV-1996; 96WC-AU00767.
 XX
 PR 30-NOV-1995; 95AU-0006911.
 PR 30-NOV-1995; 95AU-0006910.
 XX
 XX (DARA-) DARATECH PTY LTD.
 PA (PIGR-) FIG RES & DEV CORP.
 XX
 XX Hasse D, Panaccio M;
 PI
 XX WPI; 1997-310605/28.
 DR N-PSDB; AAT69201.
 XX
 PT Vaccine for treating or preventing Lawsonia intracellularis
 PT infection - especially in pigs, containing non-pathogenic form of
 PT bacterium or its components
 XX
 PS Claim 10; Page 38-42; 94pp; English.
 XX
 CC The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis
 CC can be used in vaccines to protect birds and animals against
 CC intestinal diseases, esp. to protect pigs against porcine
 CC proliferative enteropathy. It is the expression product of a DNA
 CC molecule (AAT69201) obt'd. by screening an L. intracellularis library
 CC with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)
 CC and other L. intracellularis polypeptides (AAW16680-85) can be
 CC administered as recombinant polypeptides or expressed as
 CC recombinant vaccines utilising bacterial, fungal or viral vectors.
 CC Antibodies raised against the polypeptides may be useful in
 CC immunotherapy, diagnosis of infection and detection.
 XX
 SQ Sequence 548 AA;

Alignment Scores:
 Pred. No.: 6,89e-236 Length: 548
 Score: 2708.00 Matches: 548
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 91.24% Indels: 0
 DB: 18 Gaps: 0

US-09-077-574A-1 (1-1647) x AAW16678 (1-548)

QY 1 ATGGCTCTTAAGAAATCCCTTTTCATGCTAAAGCCCGTGAAGAACTTTCACGAGTGCTA 60
 DB 1 MetAlaSerLysGluIleuPheAspAlaLysAlaArgGluLysLeuSerArgGlyVal 20
 QY 61 GATAAAGCTTGGCAATGCTGTTAAAGTAAACACTTGGACCTAAAGGCCGTAATGCTGTTAT 120
 DB 21 AspLysLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40
 QY 121 GAAAGCTTTTGGTCCCGAGTTATACAAAGATGGTGTATCTGTTGCAAGAAATTT 180
 DB 41 GluLysSerPheGlySerProValIleThrLysAspGlyValSerValAlaLysGluIle 60
 QY 181 GAACTTGAAGATAGCTTTGAAATAATAGGCGCTCAATGGTTTAAAGAAAGTAGCTCCCAAA 240
 DB 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaProLys 80
 QY 241 ACTAGCGATATGCTGTGATGGAACTACACAGCAACAGCTCCTTGGCAAGCTATTAT 300
 DB 81 ThrSerAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGluAlaIleTyr 100
 QY 301 CGTGAAGTGTAAACTGTAGCAGCTGCTGTAATCCTATGCGCATTAACAGCTGCATA 360
 DB 101 ArgGluGlyValLysLeuValAlaGalaGlyArgAsnProMetAlaIleLysArgGlyIle 120
 QY 361 GATAAAGCTGTGTTGCTTACTAAAGAACTAAGCGACATTTACAAAGCTACTCGTGAC 420
 DB 121 AspLysAlaValAlaValThrLysGluLeuSerAspIleThrLysProThrArgAsp 140
 QY 421 CMAAAGAAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGATACAAATAGGTAAT 480

141 GlnLysGluIleAlaGlnValGlyThrIleSerAlaAsnSerAspThrThrIleGlyAsn 160
 QY 481 ATCATAGCTGAAGCTATGCTAAAGTTGGAAAAGAGAGGTGTTATCACAGTTGAGAGCT 540
 DB 161 IleIleAlaGluAlaMetAlaLysValGlyLysGlyValIleThrValGluGluAla 180
 QY 541 AAAGGCTTGAACCTACATAGCTGTTGAGGAAGTGAAGTTCAGCCGTGGCTACCTC 600
 DB 181 LysGlyLeuGluThrThrLeuAspValValGluGlyMetLysPheAspArgGlyTyrLeu 200
 QY 601 TCTCCATATCTTGTAACTAATCCTGAGAAAATGGTTTGTGAACCTTGATAACCTTATATC 660
 DB 201 SerProTyrPheValThrAsnProGluLysMetValCysGluLeuAspAsnProTyrIle 220
 QY 661 CTTTGTATGAGAAAAGATTTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
 DB 221 LeuCysAsnGluLysLysIleThrSerMetLysAspMetLeuProIleLeuGlnVal 240
 QY 721 GCTAAAGTAAACCGTCCACTCCTTATTATTCGTGAAGACGTGAAGGTGAAGCACTTCCA 780
 DB 241 AlaLysValAsnArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
 QY 781 ACACCTGTAGTCAATAAGCTCCGTGAGCACTCCCAAGTTGTAGCCGTAAAAAGCTCTGCT 840
 DB 261 ThrLeuValValAsnLysLeuArgGlyAlaLeuGlnValAlaValAlaValLysAlaProGly 280
 QY 841 TTTGTGTAAGCCGTAAGCTTANGCTTGAAGATATTGCTATCTTACTGAGAGAGAACA 900
 DB 281 PheGlyGluArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluAla 300
 QY 901 ATATTTGAAGATCGTGTATAAAGCTTGAAGTGAAGCTTCTCTTCTTAGGACAGCT 960
 DB 301 IlePheGluAspArgGlyIleLysLeuGluAsnValSerLeuSerLeuGlyThrAla 320
 QY 961 AAACGTGTAGTTATTGACAAAATAATCTACTATCTGTGTAGTGTGCTGGAATAACAGAA 1020
 DB 321 LysArgValValIleAspLysGluAsnThrThrIleValAspGlyAlaGlyLysSerGlu 340
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTAAGAAACAGCTCAGATTAT 1080
 DB 341 AspIleLysAlaArgValLysGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360
 QY 1081 GATCGTGAAGAACTTCAAGAAACGCTTCGCAAACTTGTGTTGGAGTAGCTGTTATCCAT 1140
 DB 361 AspArgGluLysLeuGlnGluArgGluAlaLysLeuValGlyValAlaValIleHis 380
 QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGAGATCGTGTAGAAGATCCTCTA 1200
 DB 381 ValGlyAlaAlaThrGluThrGluMetLysGluLysLysAspArgValGluAspAlaLeu 400
 QY 1201 AATGCAACAGAGCTGCGGTTGAAGAGCTATTGTCCTGCTGGTGGTGGTACTGCTTTGTC 1260
 DB 401 AsnAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyThrAlaPheVal 420
 QY 1261 CGCTCCATTAAAGTCTTGATGATATAAACTGCTGTATGATGATGAACTTCTGCTGCACTT 1320
 DB 421 ArgSerIleLysValLeuAspIleLysProAlaAspAspGluLeuAlaGlyLeu 440
 QY 1321 AATATCATCCGTCGTTCTTTGAAGAGCTTTAGTCAAAATTTAGTCAAAATGCTGGCTAT 1380
 DB 441 AsnIleIleArgSerLeuGluProLeuArgGlnIleAlaAlaAsnAlaGlyTyr 460
 QY 1381 GAAAGCTTATTTGTAGAAAAGCTTCGTCAACCAAGATGCTTTGGATTATTAATGCT 1440
 DB 461 GluGlySerIleValValGluLysValArgGluProLysAspGlyPheGlyPheAsnAla 480
 QY 1441 GCATCAGAGAAATATGAAGCTTATTAAAGCTGTGTGCTCATTTGATCTTAAAGAAAGTTACA 1500
 DB 481 AlaserGlyGluTyrGluAspLeuIleLysAlaGlyValIleAspProLysLysValThr 500
 QY 1501 CGTATTGCAATACAAATGACAGCTACGTACCTCTTACTTCTTAACCTACAGAAATGCGCT 1560

Db 501 ArgileAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuLeuThrThrGluCysAla 520
 QY 1561 ATTCCTGAAAACACGACCTAAAGATATCCCTATGCTGGCGGTGTATGGTGGT 1620
 Db 521 IleAlaGluProGluProLysLysAspMetProMetProGlyGlyMetGlyGly 540
 QY 1621 ATGGGTGGTATGACGCTATGTAC 1644
 Db 541 MetGlyGlyMetAspGlyMetIle 548

RESULT 2

AAV23930

ID AAY23930 standard; Protein; 545 AA.

XX AC AAY23930;

XX 22-SBP-1999 (first entry)

XX Consensus amino acid sequence of a heat shock protein.

XX Heat shock protein; Hsp; immune response; immunological carrier;

XX cancer control; tumour; sarcoma; cancer; gene therapy.

XX Synthetic.

XX WO9935270-A1.

XX 15-JUL-1999.

XX 29-DEC-1998; 98WO-CA01203.

XX 31-DEC-1997; 97US-0001737.

XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

XX Mizzen L, Wisniewski J;

XX WPI; 1999-430397/36.

XX New nucleic acid encoding heat shock protein-60 from Streptococcus,

XX useful in vaccines, as carriers for other immunogens, as anticancer

XX agents and for diagnosis

XX Disclosure; Fig 10A-E; 176pp; English.

XX AAY23905-30 represent heat shock proteins (Hsps). The specification

XX describes Streptococcal Hsps, designated Hsp60. These proteins, their

XX fragments, variants and fusion proteins, are used to elicit or enhance

XX an immune response against Streptococcus, and to elicit a similar

XX response to a target antigen fused to the protein. Unlike other

XX immunological carriers, Hsp60 proteins are not immunosuppressive so

XX provide an increased response to any conjugated or fused antigen. Also,

XX where used for cancer control, they lack the side effects associated

XX with endotoxins. They can also be used to detect specific antibodies

XX and in treatment or prevention of tumours (e.g. sarcoma or cancers of

XX breast, ovary, prostate, lung, pancreas or liver). The Hsp60

XX polynucleotide is used for recombinant production of the protein, as

XX a source of primers and probes for detecting streptococci in standard

XX hybridization/amplification assays, and therapeutically in gene

XX therapy vectors.

XX SQ Sequence 545 AA;

XX Alignment Scores:

XX Pred. No.: 1.19e-159 Length: 545

XX Score: 1864.50 Matches: 380

XX Percent Similarity: 83.15% Conservative: 74

XX Best Local Similarity: 69.60% Mismatches: 85

XX Query Match: 62.82% Indels: 7

XX DB: 20 Gaps: 5

XX US-09-077-574A-1 (1-1647) x AAY23930 (1-545)

QY 7 TCTAAAGAAATCTCTTTGATGCTAAAGCCGCGTAAAGAACTTTTACGAGGTGTAGATAAA 66
 Db 2 AlaLysAspIleLysPheGlyGluAlaAaArgAlaMetLeuArgGlyValAsnAla 21
 QY 67 CTTCGAAATGCTGTTAAAGTAACACCTGGACCTTAAAGCGCGTAAATGTCGTATTGAAAG 126
 Db 22 LeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41
 QY 127 TCTTTTGGTTCCTCCAGTATTACAAAGATGGTGTATCTGTCGCAAGAAATTTGAACTT 186
 Db 42 SerPheGlyAlaProThrIleThrLysAspGlyValThrValAlaLysGluIleGluLeu 61
 QY 187 GAAGATAAGTTTCAGAAATATGGCGCTCAATGTTAAAGAGTAGCTCCCAAACTAGC 246
 Db 62 GluAspLysPheGluAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAsn 81
 QY 247 GATATTGCTGGTATGAACTACCAACAGCAACAGTCTCTTGCAACAGCTATTATTCGTGAA 306
 Db 82 AspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleValLysGlu 101
 QY 307 GGTGTAAGACTTGTACAGCTGCTGTAATCTTATGCCATTAAACGTGCATAGATAAA 366
 Db 102 GlyLeuLysAsnValAlaAlaGlyAlaAsnProMetAspLeuArgGlyIleAspLys 121
 QY 367 GCTGTTGCTGCTTACTTAAGAACTAAGCGCATTACAAAGCTACTCTGACCAAAA 426
 Db 122 AlaValAspAlaValGluGluLeuLysAlaIleAlaLysProValGluThrLysGlu 141
 QY 427 GAAATAGCTCAAGTTGGACCATTTCTGCAAACTCTGATACAACAATAGTAATATCAT 486
 Db 142 GluIleAlaGlnValAlaThrIleSerAlaAsnGlyAspGluGluIleGlyGluLeu 161
 QY 487 GCTGAAGCTATGCTTAAAGTTGAAAAGAGGTGTTATCACAGTTGAGGAGCTAAAGT 546
 Db 162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGlyThr 181
 QY 547 CTGAAACTACATTAGATGTGTTGAAGGAATGAAGTTTACCGTGTGCTACCTCTCCA 606
 Db 182 LeuGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyIleSerPro 201
 QY 607 TACTTTGTAATAATCCTGAGAAATGGTTTGTGAACCTTGATTAACCTTATATCTTTGT 666
 Db 202 TyrPheIleThrAspSerGluLysGlnLysAlaGluLeuGluAspProLeuIleLeu 221
 QY 667 AATGAGAAAGATTACTACATCAAGACATGCTTACCATTCTAGAACAGTGTCTAAA 726
 Db 222 ThrAspLysIleSerAsnIleGlnAspLeuProValLeuGluGluValAlaGln 241
 QY 727 GTAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCATTGCAACACT 786
 Db 242 AlaGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 261
 QY 787 GTAGTCAATTAAGCTCCGTGGAGCAGCTCAAGTTGTAGCCGTAAAGCTCCTGTTTGGT 846
 Db 262 ValValAsnLysLeuArgGlyThrLeuLysValValAlaValLysAlaProGlyPheGly 281
 QY 847 GAACGCGTAAAGCTATGTTGAAGATTTGCTTACTTACTGAGAGAGCAATATTT 906
 Db 282 AspArgGlyAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyGlyValIleSer 301
 QY 907 GAAGATCGTGTATTAAGCTTGAAGTGTGCTTCTTTTAGCAACAGCTAAACGT 966
 Db 302 GluGluLeuGlyLeuSerLeuGluAspAlaThrLeuGluAspLeuGluAlaLysLys 321
 QY 967 GTAGTTATTACAAAGAAATATCTTATCTGTTGATGCTGCTGGAATATAGAGATATT 1026
 Db 322 ValValValThrLysAspThrThrIleValAspGlyAlaGlyAspAlaAa--Ile 340
 QY 1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGACAAACAGCTCAGATTATGTCGT 1086
 Db 341 AlaGlyArgValAlaGlnIleArgSerGlnIleGluGluSerThrSerAspTyrAspLys 360
 QY 1087 GAAAACTTCAAGAACGCTCTTGTGCAAACTTGTGCTGAGTAGCTGTATCCATGTTGGA 1146

Db 361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 380
 QY 1147 GCTGCTACTGAACCTGAATGAAGAGAGAGATCGTGTAGAGACGCTCTAAATGCA 1206
 Db 381 AlaAlaThrGluValGluLeuLysGluArgLysAspArgValGluAspAlaLeuAsnAla 400
 QY 1207 ACAAGAGCTCGGTTGAAGAAGATTTGTCCTCGTGGTGGTACTGCTTTGTCGCTCC 1266
 Db 401 ThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuLeuArgAla 420
 QY 1267 ATTAAGCTCTGTGATTAATTAACCTGCTGTGATGATGATGAACCTGCTGGACTTAATATC 1326
 Db 421 AlaProAlaLeuAspLysLeuLysThrGluAsnGlyAspGluAlaThrGlyValAsnIle 440
 QY 1327 ATCCGCTGCTCTCTGAAGAGCTTTAGTCAAAATGCTGCAAAATGCTGCAAAATGCTGCAAGGT 1386
 Db 441 ValLeuArgAlaLeuGluAlaProLeuArgGlnIleAlaGluAsnAlaGlyLeuGluGly 460
 QY 1387 TCTATTGTTGTAAGAAAAGTTCGTGAACCAAAAGATGGTTTGGATTAAATGCTGCATCA 1446
 Db 461 SerValValGluLysValLysAsnSerGluAlaGly--GlyTyAsnAlaAlaThr 479
 QY 1447 GGAAATATGAAGACTTATTAAGCTGTGTCATGTATGATCTTAAGAAAGTTACAGCTATT 1506
 Db 480 GlyGluTyValAspMetIleAlaAlaGlyIleIleAspProThrLysValThrArgSer 499
 QY 1507 GCATTACAAAAGTCAGCATCAGTACGCTCTTACTTCTTAACCTACAGATGCGCTATTGCT 1566
 Db 500 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrGluAlaValVal 519
 QY 1567 GAAAAACACAGAACCTTAAAAAGATATGCTCT-----ATGCTTGGC-----GGTGGTATG 1614
 Db 520 AspLysProGlu--LysGluAlaAlaProAlaGlyMetProGlyMetMetGlyGlyMet 538
 QY 1615 GGTGTATGCGGTGGTATG 1632
 Db 539 GlyGlyMetGlyGlyMet 544

RESULT 3

AAY23909

ID AAY23909 standard; Protein; 549 AA.

XX AC

XX AAY23909;

XX DT 22-SEP-1999 (first entry)

XX DE Amino acid sequence of a heat shock protein.

XX KW Heat shock protein; Hsp; immune response; immunological carrier;

XX KW cancer control; tumour; sarcoma; cancer; gene therapy.

XX OS Legionella pneumophila.

XX XX WO9935270-A1.

XX PN 15-JUL-1999.

XX PF 29-DEC-1998; 98WO-CA01203.

XX PR 31-DEC-1997; 97US-0001737.

XX PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

XX PI Mizzen L, Wisniewski J;

XX XX WPI; 1999-430397/36.

XX PT New nucleic acid encoding heat shock protein-60 from Streptococcus,

XX PT useful in vaccines, as carriers for other immunogens, as anticancer

XX PT agents and for diagnosis

XX PS Disclosure; Fig 10A-E; 176pp; English.

XX AAY23905-30 represent heat shock proteins (Hsps). The specification
 CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
 CC fragments, variants and fusion proteins, are used to elicit or enhance
 CC an immune response against Streptococcus, and to elicit a similar
 CC response to a target antigen fused to the protein. Unlike other
 CC immunological carriers, Hsp60 proteins are not immunosuppressive so
 CC provide an increased response to any conjugated or fused antigen. Also,
 CC where used for cancer control, they lack the side effects associated
 CC with endotoxins. They can also be used to detect specific antibodies
 CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
 CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
 CC polynucleotide is used for recombinant production of the protein, as
 CC a source of primers and probes for detecting streptococci in standard
 CC hybridization/amplification assays, and therapeutically in gene
 CC therapy vectors.

XX SQ Sequence 549 AA;

Alignment Scores:

Pred. No.: 3,7e-158 Length: 549
 Score: 1848.00 Matches: 367
 Percent Similarity: 82.02% Conservative: 80
 Best Local Similarity: 67.34% Mismatches: 96
 Query Match: 62.26% Indels: 2
 DB: 20 Gaps: 1

US-09-077-574A-1 (1-1647) x AAY23909 (1-549)

QY 7 TCTAAGAAATCTTTTGTAGTCTAAAGCCCGTGAAGAACTTTCACGAGTGTAGATAAA 66
 Db 2 AlaLysGluLeuArgPheGlyAspAlaArgLeuGlnMetLeuAlaGlyValAla 21
 QY 67 CTTCGAAATGCTGTTAAAGTAACTTGGACCTTAAGCCGTAAGTCTGTTATTGAAAG 126
 Db 22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41
 QY 127 TCTTTTGGTTCCTCCAGATTATACAAAAGATGGTGTATCTGTGCAAAAGAAATTCGACT 186
 Db 42 SerTyGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61
 QY 187 GAAGATAAGTTGAAATATATGGCGCTCAATGGTTAAAGAGTGTCTCCCAAACTAGC 246
 Db 62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81
 QY 247 GATATTGCTGTGATGGAACCTACACAGCAGTCCCTGTGCAACAGCTATTTATCGTGA 306
 Db 82 AspThrAlaGlyAspGlyThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101
 QY 307 GGTGTAAACTTGTAGCAGTGGTCTGTAATCTATGGCCATTAAACGTGGCATAGATAAA 366
 Db 102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121
 QY 367 GCTGTTGTGCTGTTACTAAAGAACTAAGCAGCATATACAAAGCCTACTCGTGACCAAAA 426
 Db 122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141
 QY 427 GAAATAGCTCAAGTTGGACCACTTCTGCAACTCTGATACACATAGTATATCATATA 486
 Db 142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 161
 QY 487 GCTGAAGCTATGCTTAAAGTGTGAAAAGGAGGTGTATTATCACAGTTGAGGACCTAAAGT 546
 Db 162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181
 QY 547 CTTGAACTACATTAGATGTGTTCAAGAAATGAAGTGTGACCGTGGCTACTCTCTCCA 606
 Db 182 LeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyIleSerPro 201
 QY 607 TACTTTGTAACTTAATCCTCGAGAAATGTTTGTGAACTTGATTAACCTTATATCTTTGT 666
 Db 202 TyrPheIleAsnAsnGlnGlnMetSerCysGluLeuGluHisProPheIleLeuLeu 221

```
QY 667 AATGAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGACAAGTTGCTAAA 726
Db 222 ValAspLysValSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241
QY 727 GTAACCGCTCACTCCCTTATTATGCTGAAGAGCTAGAAGGTGAAGCACTTGCAACACT 786
Db 242 SerGlyArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 261
QY 787 GTAGTCAATTAAGCTCGGTGAGCAGCTCCAAAGTTGTAGCCGTAAAGCTCTGCTTTGGT 846
Db 262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281
QY 847 GAACGCGTAAAGCTATGCTTGAAGATATTGCTATCTTACTGAGAGAGCAATATTT 906
Db 282 AspArgLysAlaMetLeuGlnAspIleAlaIleuThrLysGlyGlnValIleSer 301
QY 907 GAAGATCGTGTATATAAGCTTGAAGCTTGAAGCTTGTCTTTTAGGAACAGCTAAACGT 966
Db 302 GluLuleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321
QY 967 GTAGTTATGACAAAGAAATACTATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1026
Db 322 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 341
QY 1027 AAGCTCGAGTTAAACAAATTCGTGCAAAATTCAGAAACAGCTCAGATTTATGCTGT 1086
Db 342 AsnAlaArgIleThrGlnIleArgAlaGlnMetGluThrThrSerAspTyrAspArg 361
QY 1087 GAAAAACCTCAAGAACCTCTTGTGCAAACTTGTGCTGAGTAGCTGTATCCATGTTGGA 1146
Db 362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381
QY 1147 GCTGCTACTCAAACTGAATGAAGAGAGAGATGCTGTAGAGATGCTCTTAATGCA 1206
Db 382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisaLa 401
QY 1207 ACAGAGCTCGCTGAGAGAGTATTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1266
Db 402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyValAlaLeuIleArgAla 421
QY 1267 ATTAAGCTCTGATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGATGAT 1326
Db 422 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 441
QY 1327 ATCCGCTGCTCTGTGAAGACCTTTAGCTCAATGCTGCAATGCTGCTGCTGCTGCTGCT 1386
Db 442 LeuArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 461
QY 1387 TCTATTGCTGAGAAAAGTTCGTGAACCAAGATGCTTTTGGATTTAATGCTGCATCA 1446
Db 462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481
QY 1447 GGAGATATGAAGACCTTATTAAAGCTGGTGTCAATGATCCTAATAAAGTTTACAGTATT 1506
Db 482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501
QY 1507 GCATTCAAAATGAGATCAGTACCTCTTACTTCTAATCAGATGCTGCTGCTGCTGCT 1566
Db 502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521
QY 1567 GAAAAACCAAGACCTTAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1626
Db 522 AspLeu-----ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539
QY 1627 GGTATGACCGGTATG 1641
Db 540 GlyMetGlyGlyMet 544
```

RESULT 4
AAB69060
ID AAB69060 standard; Protein; 547 AA.
XX
AC AAB69060;

```
XX 18-APR-2001 (first entry)
DT Pseudomonas aeruginosa groEL protein sequence.
XX
XX Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
KW diagnosis; detection; infection; immune response.
XX
XX Pseudomonas aeruginosa.
OS
XX WO200102577-A1.
PN
XX 11-JAN-2001.
PD
XX 03-JUL-2000; 2000WO-GB02554.
PF
XX 01-JUL-1999; 99GB-0015419.
PR (PROV-) PROVALIS UK LTD.
PA
XX Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PJH, Wilkinson MC;
XX WPI; 2001-080988/09.
XX N-PSDB; AAF32452.
XX
XX Antigenic Pseudomonas aeruginosa proteins, useful in the detection
XX and/or diagnosis of P. aeruginosa infections and for producing vaccines
XX against P. aeruginosa -
XX
XX Claim 3; Fig 25; 129pp; English.
XX
XX The present invention describes antigenic Pseudomonas aeruginosa
XX proteins (Pl). The P. aeruginosa proteins have antibacterial activity
XX and can be used in vaccines and as antagonists. The proteins or their
XX fragments, or antibodies are useful in the detection and/or diagnosis
XX of P. aeruginosa. They are also useful for producing a vaccine and
XX inducing an immune response against P. aeruginosa infection. An agent
XX capable of antagonising, inhibiting or otherwise interfering with the
XX function or expression of Pl are useful in the manufacture of a
XX medicament for the treatment or prophylaxis of P. aeruginosa infections.
XX The present sequence represents a specifically claimed P. aeruginosa
XX groEL protein sequence from the present invention.
XX
XX Sequence 547 AA;
XX
XX Alignment Scores:
XX Pred. No.: 3,99e-156 Length: 547
XX Score: 1825.50 Matches: 367
XX Percent Similarity: 80.51% Conservative: 75
XX Best Local Similarity: 66.85% Mismatches: 102
XX Query Match: 61.51% Indels: 5
XX DB: 22 Gaps: 2
XX
XX US-09-077-574A-1 (1-1647) x AAB69060 (1-547)
QY 1 ATGCTCTTAAAGAAATCCCTTTTGTGCTAAAGCCGTGAAAAAATCTTACAGAGGTGA 60
Db 1 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 20
QY 61 GATAAATCTGCAATGCTGTGTTAAAGTAAACACTTGACCTAAAGCCGTAATGCTGTATT 120
Db 21 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGTCTTTTGGTTCCTCCAGTATTACAAAGATGCTGCTATCTGTTGCAAAAGTAAT 180
Db 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACCTCAAGATAGTTTGAATATGCGCTCAATGTTTAAAGAGTAGTCCCAAA 240
Db 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 80
QY 241 ACTAGCGATATTGCTGCTGATGGAACCTACAAACAGCAACAGCTCTTGCACAAGCTATTAT 300
XX
```


Best Local Similarity: 66.61% Mismatches: 97
Query Match: 61.34% Indels: 3
DB: 16 Gaps: 2

US-09-077-574A-1 (1-1647) x AAB67381 (1-547)

QY 7 TCTAAGCAATCTTTTGATGCTAAAGCCGCGTAAAGAACTTTTCCAGAGGTAGATAAA 66
DB 2 AlAlYsGluLeuArgPheGlyAspAlaLeuGluMetLeuAlaGlyValAlaAla 21

QY 67 CTTCGAAATCTCTTAAAGTAAACACCTTGAACCTAAAGCCGCTAATGCTGTTATGAAAG 126
DB 22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41

QY 127 TCTTTGGTCCCGAGTTATTAACAAAGATGGTGATCTCTTGGCAAGAAATTTGAACCT 186
DB 42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

QY 187 GAAGATAAGTTGAATATGGCGCTCAATGTTAAAGAGTAGCTCCCAAACTAGC 246
DB 62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81

QY 247 GATATTCTCTGATGGAATCTAACACAGCAACAGTCTTCCACAAAGCTATTATTCGTGAA 306
DB 82 AspThrAlaGlyAspGlyThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101

QY 307 GGTGTAAACTGTAGCAGCTGGTGAATCTTATGCGCATTAACCTGCGCATAGATAAA 366
DB 102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

QY 367 GCTGTGTTCTCTTAAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAG 426
DB 122 AlaValLeuAlaValThrLysLysLeuAlaMetSerLysProCysLysAspSerLys 141

QY 427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACTAGGTAAATATCATA 486
DB 142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIle 161

QY 487 GCTGAAGCTATGGCTAAAGTTGGAAGAGGAGTGTATACAGTTGAGGAAGCTAAAGGT 546
DB 162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

QY 547 CTTCGAACTACATTAGATGCTGTGAAGGAATGAAGTTGACCGTGTCTCTCTCCA 606
DB 182 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

QY 607 TACTTTCTAATCTCTGAGAAATGGTTTGTGAACCTGTGATACCCCTTATATCTTTGT 666
DB 202 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 221

QY 667 AATGAGAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTTGCTAAA 726
DB 222 ValAspLysLysValSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241

QY 727 GTAAACGCTCCACTCTTATTATTGCTGAAGCTAGAGTGAAGCACTTGCAACACT 786
DB 242 SerGlyArgProLeuLeuIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 261

QY 787 GTAGTCAATAAGCTCCGCTGGAGCACTCCCAAGTTGTAGCCGTAAAGCTCTGGTTTGGT 846
DB 262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

QY 847 GAACGCCGTAAAGCTATCTTGAAGATATTCCTATCTCTTACTGGAGGAGAGCAATATT 906
DB 282 AspArgGlyLysAlaMetGluAspIleAlaIleLeuThrLysGlyGlnValIleSer 301

QY 907 GAAGATCGTGTATPAAAGCTTGAATCTGATGCTTCTTTAGGAACAGCTAAAGCT 966
DB 302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321

QY 967 GTAGTTATTGACAAAGAAATACTACTATCGTTGATGCTGGTAAATATCAGAAATATT 1026
DB 322 IleIleValThrAspGlu---ThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 340

QY 1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGACAAAGCTCAGATTTATGATCGT 1086
DB 341 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 360

QY 1087 GAAAACTTCAAGACAGCTCTTCAAAACTTGTGGTGGAGTAGCTGTATTCATGTTGGA 1146
DB 361 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 380

QY 1147 GCTGCTACTGAACTGAAATGAAGAGAGAGAGATCGCTAGAGATGCTCTAAATGCA 1206
DB 381 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 400

QY 1207 ACAAGAGCTCGCTTGAAGAGAGGTATTGCTCCCTGGTGGTGTACTGCTTTTGTCCGCTCC 1266
DB 401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyValAlaLeuIleArgAla 420

QY 1267 ATTTAAAGTCTCTGATGATATTAAACCTGCTGATGATGAATGAATGCTGCTGCTAAATC 1326
DB 421 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspGlnAsnMetGlyIleAsnIle 440

QY 1327 ATCGTCTGCTCTTGAAGAGCTTTACGTCAAATTCGCTCAATGCTGCTGCTGCTGCTGCT 1386
DB 441 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 460

QY 1387 TCTATTCTTCTAGAAAAAGTTCTGTGAACCAAAAGATGTTTGGATTATTAATGCTGCATCA 1446
DB 461 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 480

QY 1447 GGAGAATATCAAGACCTTATTAAAGCTGGTGTCTATTCATCTCTAAAGTTACACGTATT 1506
DB 481 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 500

QY 1507 GCATTCAAAATGCAGCATCAGTAGCTCTCTTCTTCTTAACTACAGATGCTGCTATTGCT 1566
DB 501 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 520

QY 1567 GAAAAACCAAGCACTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1626
DB 521 AspLeu-----ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 538

QY 1627 GGTATGACCGGTATG 1641
DB 539 GlyMetGlyGlyMet 543

RESULT 6
AAB69061 standard; Protein; 574 AA.
ID AAB69061;
AC AAB69061;
XX
XX 18-APR-2001 (first entry)
XX
XX pTrcHisB expression vector groEL fusion protein sequence.
XX
XX Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
XX
XX diagnosis; detection; infection; immune response.
XX
XX Pseudomonas aeruginosa.
XX
XX Synthetic.
XX
XX WO200102577-A1.
XX
XX 11-JAN-2001.
XX
XX 03-JUL-2000; 2000WO-GB02554.
XX
XX 01-JUL-1999; 99GB-0015419.
XX
XX (PROV-) PROVALIS UK LTD.
XX
XX Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PUH, Wilkinson MC;
XX

DR WPI; 2001-080988/09.
 XX N-PSDB; AAF32453.
 PT Antigenic Pseudomonas aeruginosa proteins, useful in the detection
 PT and/or diagnosis of P. aeruginosa infections and for producing vaccines
 PT against P. aeruginosa -
 XX
 PS
 PS Claim 3; Fig 26; 129pp; English.
 XX

CC The present invention describes antigenic Pseudomonas aeruginosa
 CC proteins (PI). The P. aeruginosa proteins have antibacterial activity
 CC and can be used in vaccines and as antagonists. The proteins or their
 CC fragments, or antibodies are useful in the detection and/or diagnosis
 CC of P. aeruginosa. They are also useful for producing a vaccine and
 CC inducing an immune response against P. aeruginosa infection. An agent
 CC capable of antagonising, inhibiting or otherwise interfering with the
 CC function or expression of PI are useful in the manufacture of a
 CC medicament for the treatment or prophylaxis of P. aeruginosa infections.
 CC The present sequence represents a specifically claimed pTcrH:SB
 CC expression vector groEL fusion protein sequence from the present
 CC invention.

XX
 SQ Sequence 574 AA;

Alignment Scores:

Pred. No.: 1-57e-155 Length: 574
 Score: 1819.00 Matches: 365
 Percent Similarity: 80.44% Conservative: 75
 Best Local Similarity: 66.73% Mismatches: 103
 Query Match: 61.29% Indels: 4
 Ds: 22 Gaps: 1

US-09-077-574A-1 (1-1647) x AAB69061 (1-574)

QY 1 ATGGCTTCTAAGAAATCTTTTGTGTAAGCCGCTGAAACCTTTCACAGGTGTA 60
 Db 31 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysMetLeuValGlyVal 50
 QY 61 GATAAATCTGCAATCTGTTAAAGTAACACTTGGACCTTAAAGCCGCTAATGCTTATT 120
 Db 51 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 70
 QY 121 GAAAGTCTTTGTTGCCAGTTATACAAAGATGGTGTATCTGTTCCAAAGAAAT 180
 Db 71 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 90
 QY 181 GAACCTTGAAGATACTTTGAAATATGGCGCTCAAACTGGTTAAAGAACTAGTCCCAA 240
 Db 91 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 110
 QY 241 ACTAGCGATATTGCTGGTGAATCAACAGCAACAGCTCTTGCACAAAGCTATTAT 300
 Db 111 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaVal 130
 QY 301 CGTGAAGGTGAAACTTGTAGCAGCTGGTCTATCTCCATCGCCATTAAGCTGGCATA 360
 Db 131 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 150
 QY 361 GATAAGCTGTTGTTGCTTACTTAAAGAACTAAGCGCATTTACAAAGCTTCTGTCAC 420
 Db 151 AspLysAlaThrValAlaIleValAlaGlnLeuLysGluLeuAlaLysProCysAlaAsp 170
 QY 421 CAAAGAAGATAGCTCAAGTTGGAACTTCTGCAAACTCTGATACAACTAGATTAAT 480
 Db 171 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyGln 190
 QY 481 ATCATAGCTGAAGCTATGCTTAAAGTGGAAAGAGGTGTTATCACAGTTGAGGAAGCT 540
 Db 191 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 210
 QY 541 AAAGTCTTGAACCTACATAGATCTGTTGTAAGGAATGAAGTTGACCGGTGCTACCTC 600
 Db 211 SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 230

QY 601 TCTCCATCTTGTAACTAATCCTGAGAAATCGTTTGTCAACTTGTATACCTTATATC 660
 Db 231 SerProTyrPheValAsnLysProAspThrMetAlaAlaGlnLeuAspSerProLeuLeu 250
 QY 661 CTTTGTAAATCAGAAAAGATTACTAGCATCAAGACATGTACCAATCTTAGAACAGTT 720
 Db 251 LeuLeuValAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 270
 QY 721 GCTAAAGTAAACCGTCCACTCTTATTATTGTGAAGACGTAGAAGGTGAAGCACTTGA 780
 Db 271 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 290
 QY 781 ACACCTGTAGTCAATAAGCTCGTGCAGCACCTCAAGTTCTAGCCCTTAAAGCTCTG 840
 Db 291 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 310
 QY 841 TTTGTGAACCGCTAAAGCTATGCTTGAAGATATTGCTTATCTTACTGAGGAGAAGCA 900
 Db 311 PheGlyAspArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 330
 QY 901 ATATTGAAGATCGTGGTATTAAGCTTGAAGATGTAAGCTTGTCTTCTTAGGAACAGCT 960
 Db 331 IleSerGluGluValGlyLeuSerLeuGluGlyAlaThrLeuGluHisLeuGlyAsnAla 350
 QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTGTTGATGCTGCTGAAATCAGAA 1020
 Db 351 LysArgValValIleAsnLysGluAsnThrThrIleLeuAspGlyAlaGlyValGlnAla 370
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGACAAATGAAAGAAACAGACTCAGATTAT 1080
 Db 371 AspIleGluAlaArgValLeuGlnIleArgLysGlnIleGluGluThrThrSerAspTyr 390
 QY 1081 GATCGTGAAGAACTTCAAGAACGTCTTGCAAAACCTTGTGTTGGTGGAGTAGCTGTATCCAT 1140
 Db 391 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleLys 410
 QY 1141 GTTGAGCTGCTACTGAACTCAATGAAAGAAAGAAAGAGTCTGCTAGAAAGTCTCTA 1200
 Db 411 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 430
 QY 1201 AATGCAACAAAGAGCTCGCGTTGAAGAAGTATTGTCCTGTTGGTGGTGTCTGTTGTC 1260
 Db 431 HisAlaThrArgAlaAlaValGluGluValValProGlyGlyValAlaLeuVal 450
 QY 1261 CGCTCCATTAAAGTCTTGATGATATTAACTGCTGATGATGATGAACCTGCTGACTT 1320
 Db 451 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluGlnAsnValGlyIle 470
 QY 1321 AATATCATCCGCTGTTCTTCTGAGAGCTTTACGTCAAAATGCTGCAAACTGCTGCTAT 1380
 Db 471 AlaLeuLeuArgArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyAsp 490
 QY 1381 GAAGTTCTTATTGTTGTAGAAAAGTTCGTGAACCAAAAGATGTTGTTTGAATTAATGCT 1440
 Db 491 GluProSerValValValAspLysValLysGlnGlySerGlyAsnTyrGlyPheAsnAla 510
 QY 1441 GCATCAGAGATATGAAGACCTTATTAAAGCTGTGTCTGATGATGATGATGATGATGAT 1500
 Db 511 AlaThrGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 530
 QY 1501 COTATTGATCAAAATGACAGCTAGCTAGCTCTTCTTACTTCTTAACCTACAGAATGCGCT 1560
 Db 531 ArgSerAlaLeuGlnAlaAlaSerIleGlyGlyLeuMetIleThrThrGluAlaMet 550
 QY 1561 ATTGCTGAAAACCCAGAACCTTAAAGATATGCTTATGCTTGGCGGTGGTATGGGTGCT 1620
 Db 551 ValAlaGluIleValGluAspLysProAlaMet-----GlyGlyMetProAsp 566
 QY 1621 ATGGTGTATCGACGGTATG 1641
 Db 567 MetGlyGlyMetGlyGlyMet 573

RESULT 7

AAV23915
ID RAY23915 standard; Protein; 545 AA.

XX AC AAY23915;

XX DT 22-SEP-1999 (first entry)

XX DE Amino acid sequence of a heat shock protein.

XX KW Heat shock protein; Hsp; immune response; immunological carrier;
cancer control; tumour; sarcoma; cancer; gene therapy.

XX OS Neisseria meningitidis.

XX PN WO9935270-A1.

XX PD 15-JUL-1999.

XX PF 29-DEC-1998; 98WO-CA01203.

XX PR 31-DEC-1997; 97US-0001737.

XX PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

XX PI Mizzen L, Wisniewski J;

XX DR WPI; 1999-430397/36.

XX PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
useful in vaccines, as carriers for other immunogens, as anticancer
agents and for diagnosis

XX PS Disclosure; Fig 10A-E; 176pp; English.

XX CC AAY23905-30 represent heat shock proteins (Hsps). The specification
describes Streptococcal Hsps, designated Hsp60. These proteins, their
fragments, variants and fusion proteins, are used to elicit or enhance
an immune response against Streptococcus, and to elicit a similar
response to a target antigen fused to the protein. Unlike other
immunological carriers, Hsp60 proteins are not immunosuppressive so
provide an increased response to any conjugated or fused antigen. Also,
where used for cancer control, they lack the side effects associated
with endotoxins. They can also be used to detect specific antibodies
and in treatment or prevention of tumours (e.g. sarcoma or cancers of
breast, ovary, prostate, lung, pancreas or liver). The Hsp60
polynucleotide is used for recombinant production of the protein, as
a source of primers and probes for detecting streptococci in standard
hybridization/amplification assays, and therapeutically in gene
therapy vectors.

XX SQ Sequence 545 AA;

Alignment Scores: 3,93e-155 Length: 545
Pred. No.: 1814.50 Matches: 363
Score: 80.88% Conservative: 77
Best Local Similarity: 66.73% Mismatches: 103
Query Match: 61.14% Indels: 1
DB: 20 Gaps: 1

US-09-077-574A-1 (1-1647) x AAY23915 (1-545)

QY 1 ATGGCTCTCAAGAATCCTTTTTCATGCTAAGACCCGTGAAACCTTTTCAGAGGTGTA 60
DB 1 MetAlaAlaLysAspValGlnPheGlyAsnGluValArgGlnLysMetValAsnGlyVal 20
QY 61 GATAAAGTTCGAAATGCTGTTAAAGTAAACACTTGACCTTAAGCGCGTAAATGCTGTTATT 120
DB 21 AsnIleLeuAlaAsnAlaValArgValThrLeuGlyProLysGlyArgAsnValValVal 40
QY 121 GAAAAGTCTTTTGGTCCCACTTATCAAAAGATGGTGATCTCTTGCAGAAAGAAATT 180

Db 41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60
QY 181 GAACCTTGAACATAAGTTTGAATAATATGGCGCTCAATGCTTAAAGAGTAGCTCCCAAA 240
Db 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTCTGCTGATGGAACCTACAAACAGCAACAGCTCCTTGCAAGCTATTAT 300
Db 81 ThrAsnAspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
QY 301 CGTGAAGGTGTAACCTTTAGCAGCTGTCGTAATCCTATGCGCATTAACCTGAGGATA 360
Db 101 AlaGluGlyMetLysTyrValThrAlaGlyMetAsnProThrAspLysAspGlyIle 120
QY 361 GATAAAGCTGCTGCTGCTTACTAAAGAACTAAGCGACATTAACAAAGCCTACTCGTGAC 420
Db 121 AspLysAlaValAlaAlaLeuValGluLeuLysAsnIleAlaLysProCysAspThr 140
QY 421 CAAAAGAAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGATACCAATAGGTAAT 480
Db 141 SerLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspGluGlnValGlyAla 160
QY 481 ATCATAGCTCAAGCTATGCTGCTAAAGTTGGAAGAGAGGTGTTATCACAGTTGAGAACT 540
Db 161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180
QY 541 AAAGCTCTTGAACACTACATTAGATGCTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db 181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATCTTTGTAACTAATCTCTGAGAAATGTTTGTGAACCTTATGATAACCTTATATC 660
Db 201 SerProTyrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220
QY 661 CTTTGTATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGACAAGTT 720
Db 221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
QY 721 GCTAAAGTAAACCGTCCACCTCTTATTATGCTCAAGACGTAGAGGTGAAGCACTTCCA 780
Db 241 AlaLysAlaSerArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATCTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAAGCTCCTGT 840
Db 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCCGCTAAGCTATGCTTGAAGATATGCTATCTCTTACTGAGGAGCAACA 900
Db 281 PheGlyAspArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 300
QY 901 ATATTTGAAGATCGTGGTATAAAGCTGAAAAATGTAAGCTTGTCTTTTAGGAACAGCT 960
Db 301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrLeuAspAspLeuGlyGlnAla 320
QY 961 AAACGTGTAGTTATTGACAAAAGAAAAATACTACTATCGTTGATGGTCTGGAATAATCAGAA 1020
Db 321 LysArgIleGluIleGlyLysGluAsnThrThrIleIleAspGlyPheGlyAspAlaAla 340
QY 1021 GATATTAAAGTCGAGTTAAACAAATTCGTGCAAAATGTAAGAAACAGCTCAGATTAT 1080
Db 341 GlnIleGluAlaArgValAlaGluIleArgGlnGlnIleGluThrAlaThrSerAspTyr 360
QY 1081 GATCGTGAACAACTTCAAGAACGCTTGCAGAACTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db 361 AspLysGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValLysLys 380
QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAAGAGAAAGAGGATCGTGTAGAAGATGCTCTTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAspArgValGluAspAlaLeu 400
QY 1201 AATGCAACAGAGCTGCGGTTGAAAGAGATATGTCCTCGTGGTGGTGGTACTGCTTTGTC 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuLeu 420

```
QY 1261 CGTCCATTAAAGTCCTTGATGATATTAACCTGCTGATGATGATGAACCTTGCTGGACTT 1320
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 421 ArgAlaArgAlaAlaLeuGluAsnLeuHisThrClyAsnAlaAspGlnAspAlaGlyVal 440
QY 1321 AATATCATCGTCTGCTCTTGTGAAGACCTTTTACGTCGAATTCGTCGAATTCGCTGAT 1380
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 441 GlnIleValLeuArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyGly 460
QY 1381 GAAGGTTCTATTCTGTAGAAAAGTTTCGTAACCAAGATGGTTTGCATTTAATGCT 1440
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 461 GluProSerValValValAsnLysValLeuGluGlyLysGlyAsnTyrglyTyrsAla 480
QY 1441 GCATCAGGAAATPATGAAGACCTTATTAAAGCTGCTGCTCATTCCTTCTTAACTACAGAAATGCT 1500
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 481 GlySerGlyGluTyrglyAspMetIleGluMetGlyValLeuAspProAlaLysValThr 500
QY 1501 CGTATTGCATTCAAAATGAGATCAGTACGATGCTGCTTCTTCTTAACTACAGAAATGCT 1560
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 501 ArgSerAlaLeuGlnHisAlaAlaSerIleAlaGlyLeuMetLeuThrThrAspCysMet 520
QY 1561 ATGCTGAAACCAACAGACCTTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 521 IleAlaGluIleProGluGluLysProAlaMetPro---AspMetGlyGlyMetGlyGly 539
QY 1621 ATGGGTGGTATG 1632
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 540 MetGlyGlyMet 543
```

RESULT 8

AAV75747
ID AAV75747 standard; Protein; 544 AA.

XX AAV75747;

XX 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 982 protein sequence SEQ ID NO:2966.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.

OS Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI: 2000-062150/05.

DR N-PSDB; AA254509.

Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics

XX

PS Claim 2; Page 1388; 1453pp; English.
XX AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA2575941
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC *Neisseria meningitidis* (e.g. meningitis and septicemia), to detect the
CC presence of *Neisseria meningitidis*, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 544 AA;

Alignment Scores: 5.96e-155 Length: 544
Pred. No.: 1812.50 Matches: 362
Score: 80.88% Conservative: 78
Percent Similarity: 66.54% Mismatches: 103
Best Local Similarity: 61.07% Indels: 1
Query Match: 21 Gaps: 1
DB:

US-09-077-574A-1 (1-1647) x AAV75747 (1-544)

QY 1 ATGGCTTCTAAAGAAATCCTTTTGTGATGCTAAAGCCCGTGAAGAAATCTTCCAGAGGTGA 60
Db 1 MetAlaAlaLysAspValGlnPheGlyAsnGluValArgGlnLysMetValAsnGlyVal 20
QY 61 GATAAATCTGCAATGCTCTTAAAGTAACTGACCTTGAAGCCGCTAATGCTGCTTATT 120
Db 21 AsnIleLeuAlaAsnAlaValArgValThrLeuGlyProLysGlyArgAsnValVal 40
QY 121 GAAAGTCTTTTGGTCTCCAGTATTACAAAAGATGCTGCTCTTCCCAAGAAAT 180
Db 41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60
QY 181 GAACCTGAAGATGAAGTTTGAAATATGGCGCTCAAAATGTTTAAAGAGTAGCTCCCAA 240
Db 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCATATTGCTGGTGCATGCACTACACAGCAACAGTCTCTGCAAGCTATTAT 300
Db 81 ThrAsnAspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
QY 301 CGTGAAGGTGTAAGCTTGTAGCAGCTGCTGCTAATCTATGCTTAAAGCTGAGCATA 360
Db 101 AlaGluGlyMetLysTyrrValThrAlaGlyMetAsnProThrAspLeuLysArgGlyIle 120
QY 361 GATAAGCTGTTGTTGCTGTTTAAAGAACTAAGCGACATTTACAAAGCTTACTCGTGAC 420
Db 121 AspLysAlaValAlaAlaLeuValGluGluLeuLysAsnIleAlaLysProCysAspThr 140
QY 421 CAAAAGAAATAGCTCAAGTTGCAACCATTTCTGCAAACTCTGATACAAATAGGTAAT 480
Db 141 SerLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspGluGlnValGlyAla 160
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTCGAAAAGGAGGTGTTTATCACAGTTGAGGAGCT 540
Db 161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180
QY 541 AAAGCTCTTGAACCTACATGATGTTGTTGAAGGAATGAAGTTTTCACCGTGGCTACCTC 600
Db 181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrrLeu 200
QY 601 TCTCCATCTTTGTAACTAATCCTCAGAAAATGTTTGTGAACTTGTAACTTATATC 660
Db 201 SerProTyrrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220
QY 661 CTTTGTATGAGAAAAGAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720

DB 221 LeuLeupheAspLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
QY 721 GCTAAAGTAAACCGTCCACCTCTTATTATGCTGAGACAGTAGAGTGAGCACTTGCA 780
DB 241 AlalyeAlaSerArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATCTGTAGTCAATAAGCTCCGTGGAGCACTCAAGTCTAGCCGTAAAGCTCTGGT 840
DB 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGCTAAAGCTATGCTTGAAGATATTGCTATCTCTACTGAGGAGAGCA 900
DB 281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 300
QY 901 ATATTTGAAGATCGTGTATAAAGCTTGAAGATTTAGCTGTCTCTTTAGGAACAGCT 960
DB 301 IleSerGluGluValGlyLeuSerIleGluLysAlaThrLeuAspAspLeuGlyGlnAla 320
QY 961 AAACGTGTAGTTTACAAAGAAATATCTACTATCTGTTGATGCTGCTGAAATCAGAA 1020
DB 321 LysArgIleGluIleGlyLysGluAsnThrIleIleAspGlyPheGlyAspAlaAla 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAGATTTAT 1080
DB 341 GlnIleGluAlaArgValAlaGluIleArgGlnGlnIleGluThrAlaThrSerAspTyr 360
QY 1081 GATCGTGAACAACTTCAAGACGCTTGTGAAAATCTGTTGGTGGAGTAGCTTTATCAT 1140
DB 361 AsplysGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTGGAGCTCTACTGAACTGAATCAAGAGAGAGAGATCGTGTAGAGTGTCTCTA 1200
DB 381 ValGlyAlaAlaThrGluValGluMetLysGluLysAspArgValGluAspAlaLeu 400
QY 1201 AATGCAACAAAGAGCTGGGTGTAAGAGGATTTGTCCTGCTGCTGCTGCTTTGTC 1260
DB 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuLeu 420
QY 1261 CGTCCATTAAGCTCCTGTAGATATTAAACCTGCTGATGATGATGATGATGATGAT 1320
DB 421 ArgAlaArgAlaAlaLeuGluAsnLeuHisThrGlyAsnAlaAspGlnAspAlaGlyVal 440
QY 1321 AATATCATCGTCTCTCTTGTAGAGACCTTTAGTCAATTTGCTGCAATGCTGCTAT 1380
DB 441 GlnIleValLeuArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyGly 460
QY 1381 GAAGGTTCTATTGTTGAGAAAAGTTTCGTGAACCAAAAGATGTTTGGATTATATGCT 1440
DB 461 GluProSerValValValAsnLysValLeuGluGlyLysGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGAGATATGAGACCTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 481 GlySerGlyGluTyrGlyAspMetIleGluMetGlyValLeuAspProAlaLysValThr 500
QY 1501 CGTATTGCATTACAAATGAGATCAGATCAGCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
DB 501 ArgSerAlaLeuGlnHisAlaAlaSerIleAlaGlyLeuMetLeuThrThrAspCysMet 520
QY 1561 ATTGCTCAAAACCAAGAACCTTAAAGATATGCTTATGCTGCTGCTGCTGCTGCTGCT 1620
DB 521 IleAlaGluIleProGluAspLysProAlaMetPro---AspMetGlyGlyMetGlyGly 539
QY 1621 ATGGGTGGTATG 1632
DB 540 MetGlyGlyMet 543
RESULT 9
ABP77245
ID ABP77245 standard; Protein; 544 AA.
XX
AC ABP77245;
XX

DT 07-MAR-2003 (first entry)
XX N. gonorrhoeae amino acid sequence SEQ ID 1020.
DE Antibacterial; infection; vaccine; gene therapy.
KW Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX 12-FEB-2002; 2002WO-IB02069.
XX 12-FEB-2001; 2001GB-0003424.
(CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Massignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ38215.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 262; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention.
XX Sequence 544 AA;
Alignment Scores:
Pred. No.: 2,56e-154 Length: 544
Score: 1805.50 Matches: 361
Percent Similarity: 80.70% Conservative: 78
Best Local Similarity: 66.36% Mismatches: 104
Query Match: 60.83% Indels: 1
DB: 24 Gaps: 1
US-09-077-574A-1 (1-1647) x ABP77245 (1-544)
QY 1 ATGCTTCTTAAAGAAATCTCTTTTGTATGCTTAAAGCCGCTGAAAAAATTTTCCAGAGGTGA 60
DB 1 MetAlaAlaLysAspValGlnPheGlyAsnGluValArgGlnLysMetValAsnGlyVal 20
QY 61 GATAAACTTCCAAATGCTGTTAAAGTAACACTTGGACCTAAAGCCGCTTAATGCTGTTATT 120
DB 21 AsnIleLeuAlaAsnAlaValArgValThrLeuGlyProLysGlyArgAsnValVal 40
QY 121 GAAAGTCTTTTGGTCCCGAGTTATTACAAAGATGCTGATCTGTTGCAAGAAATTT 180
DB 41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60
QY 181 GAATCTCAAGATAGTTTGAATAATATGCGCTCAATGTTAAAGAGTAGCTCCCAAA 240
DB 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCATATTGCTGCTGATGGAACATACACAGCAACAGTCTCTTGCACAGCTATTATT 300
DB 81 ThrAsnAspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
QY 301 CGTGAAGGTGTAACCTTTGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 101 AlaGluGlyMetLysTyrValThrAlaGlyMetAsnProThrAspLeuLysArgGlyIle 120

PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-PSDB; AA254507.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX Claim 2; Page 1385; 1453pp; English.
 XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254577 to AA254579 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisserial bacteria* (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria bacteria*, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 544 AA;

Alignment Scores: 2,39e-149 Length: 544
 Pred. No.: 1750,50 Matches: 353
 Score: 79,23% Conservative: 78
 Percent Similarity: 64,89% Mismatches: 112
 Best Local Similarity: 58,98% Indels: 1
 Query Match: 21 Gaps: 1
 DB: 1

US-09-077-574A-1 (1-1647) x AA275745 (1-544)

QY 1 ATGGCTTCTAAGAAATCCTTTTGATGCTAAAGCCCGTGAAGAACTTTCAGAGGTGA 60
 Db 1 IleAlaSerGluAsnLeuArgPheAspAsnArgPheLeuGlnLysMetValAsnGlyVal 20
 QY 61 GATAAAGTGCCTTAAAGTAACTGACCTTGGACCTTAAAGCGCGTAATCGCTATT 120
 Db 21 AsnIleLeuProAlaAlaAspTrpValAlaLeuGlyAlaLysGlyArgAsnValValVal 40
 QY 121 GAAAGTCTTTGGTTCCCGCTATTACAAAGATGGTGTATCTGCTCAAGAAATTT 180
 Db 41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60
 QY 181 GAAGTGAAGATAAGTTTGAAATATGGCGCTCAAAATGGTTAAAGAGTAGCTCCCAA 240
 Db 61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
 QY 241 ACTACGATATGCTGGTGGTGAAGTAACTACAGCAAGCAAGCTCTTGCACAGCTATTAT 300
 Db 81 ThrAsnAspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
 QY 301 CGTGAAGTGTAAACTTGTAGACAGCTGCTGCTAATCTATGCCATTAAAGCTGCATA 360
 Db 101 AlaGluGlyMetLysTyrValThrAlaGlyMetAsnProThrAspLeuLysArgGlyIle 120
 QY 361 GATTAAGCTGTTGTTGCTTACTACTAAGCAACTAACGACATTACAAAGCTACTCGTAC 420
 Db 121 AspLysAlaValAlaAlaLeuValGluLeuLysAsnIleAlaLysProCysAspThr 140
 QY 421 CAAAGAAATAGCTCAAGTGGACCATTTCTGCAAACTCTGATACAACTAGTAAAT 480
 Db 141 SerLysGluIleAlaGlnValGlySerIleAlaAsnSerAspGluGlnValGlyAla 160
 QY 481 ATCATAGCTGAAGCTATGCTAAAGTGGAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
 Db 161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

QY 541 AAAGGCTTGGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
 Db 181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
 QY 601 TCTCCATCTTTGTAATAATCTCTGAGAAATGTTTGTGAACCTTGAATACCTCTTATTC 660
 Db 201 SerProTyrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220
 QY 661 CTTTGTAAATGAGAAAGAAATTTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
 Db 221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
 QY 721 GCTAAAGTAAACCGCTCCACTCTCTTATTGCTGAAGACCTAGACGTGAAGCAGCTGCA 780
 Db 241 AlaLysAlaSerArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
 QY 781 ACATTTGTAGTCAATAAGCTCCGTGGAGCAGCTCCAAAGTTGTAGCCGTAAAGCTCCCTGGT 840
 Db 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280
 QY 841 TTTGCTGAAGCGCGTAAAGCTATGCTTGAAGATTTGCTTACTTCTTACTGAGGAGAGCA 900
 Db 281 PheGlyAspArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyValVal 300
 QY 901 ATATTTGAAGATCGTGGTATATAAGCTTGAAGCTTGAAGCTTGTCTTCTTTAGCAAGCT 960
 Db 301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrLeuAspLeuGlyGlnThr 320
 QY 961 AAACGTGTAGTTTACAAAGAAATATCTATCTATCTGTTGATGCTGCTGGAATAACAGAA 1020
 Db 321 LysArgIleGluIleGlyGluAsnThrThrValIleAspGlyPheGlyAspAlaAla 340
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACACAGCTCAGATTAT 1080
 Db 341 GlnIleGluAlaArgValAlaGluIleArgGlnGlnIleGluThrAlaThrSerAspTyr 360
 QY 1081 GATCGTGAAGAACTTCAAGACGCTCTTGCAGAACTTGTGGTGGAGTAGCTGTATCCCAT 1140
 Db 361 AspLysGluLysLeuGlnGluArgValAlaLysLeuAlaGlyValAlaValIleLys 380
 QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAGAGAGAGAGGATCGTGTAGAAGATGCTCTTA 1200
 Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAspArgValGluAspAlaLeu 400
 QY 1201 AATGCAACAGAGCTCGCTGCAAGAGGTATTGCTCCCTGGTGGTGGTACTGCTTTGTC 1260
 Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuLeu 420
 QY 1261 CGCTCCATTAAGTCTCTTATGATATATAACCTGCTGATGATGATGATGATGATGATGAT 1320
 Db 421 ArgAlaArgAlaAlaLeuGluAsnLeuHisThrGlyAsnAlaAspGlnAspAlaGlyVal 440
 QY 1321 AATATCATCCGCTGCTTCTTGAAGAGCTTTTACCTCAAAATTTGCTGCAAAATGCTGGCTAT 1380
 Db 441 GlnIleValLeuArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyGly 460
 QY 1381 GAAGGCTTCTATTGTTGTAAGAACTTCGTGACCAACCAAGATGGTTTGGATTTATGCT 1440
 Db 461 GluProSerValValValAsnLysValLeuGluLysGlyAsnTyrGlyTyrAsnAla 480
 QY 1441 GCATCAGAGAGATATGAAGACCTTATTAAAGCTGGTGTCTTATGATGATGATGATGATGAT 1500
 Db 481 GlySerGlyGluTyrGlyAspMetIleGlyMetGlyValLeuAspProAlaLysValThr 500
 QY 1501 CGTATTGATTAATAAATGCAGCATCAGTAGCTCTCTTACTTCTTACTTCTTACTTCTTACT 1560
 Db 501 ArgSerAlaLeuGlnHisAlaSerIleAlaGlyLeuMetLeuThrThrAspCysMet 520
 QY 1561 ATGCTGAAAAACAGACCTTAAAAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 521 IleAlaGluIleProGluGluLysProAlaPro---AspMetGlyGlyMetGlyGly 539
 QY 1621 ATGGTGTGTATG 1632

```

Db      540 MetGlyGlyMet 543
|||||
RESULT 12
ABB99014
ID ABB99014 standard; protein; 545 AA.
XX AC ABB99014;
XX DT 12-DEC-2002 (first entry)
XX DE Detrimental organism controller protein #6.
XX KW Detrimental organism; enterobacter.
XX OS Enterobacter aerogenes.
XX PN JP2002223764-A.
XX PD 13-AUG-2002.
XX PF 31-JAN-2001; 2001JP-0023255.
XX PR 31-JAN-2001; 2001JP-0023255.
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX WP I; 2002-694867/75.
XX DR N-PSDB; ABQ81658.
XX PT A novel detrimental organism controller for limiting detrimental
XX PT organisms -
XX PS Claim 8; Page 41-43; 78pp; Japanese.
XX CC The invention relates to a protein that acts as a detrimental organism
XX CC controller and is isolated from Enterobacter aerogenes. The invention
XX CC encompasses a detrimental organism-resistant plant producing the protein
XX CC of the invention in an effective amount for controlling detrimental
XX CC organisms. A composition of the invention is useful for the preparation
XX CC of a detrimental organism controller. The current sequence represents a
XX CC detrimental organism controller protein of the invention.
XX SQ Sequence 545 AA;

Alignment Scores:
Pred. No.: 3,27e-149 Length: 545
Score: 1749.00 Matches: 354
Percent Similarity: 78.21% Conservative: 73
Best Local Similarity: 64.84% Mismatches: 115
Query Match: 58.93% Indels: 4
DB: 23 Gaps: 2

US-09-077-574A-1 (1-1647) x ABB99014 (1-545)
QY 1 ATGGCTTCTTAAGAAATCCCTTTTGTATGCTAAAGCCCGTGAAGAACTTTCAGAGTGTA 60
Db 1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20
QY 61 GATAAAGCTTGCAGATGCTGTTAAAGTAACACTTGGACCTTAAGGCCGTAATGTCGTTATT 120
Db 21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGAGCTTTTGGTCCCGAGTTATTACAAAGATGGTGTATCTGTTGTCGAAAGAAATT 180
Db 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluile 60
QY 181 GAACCTTGAAGTAAGTTTGAATAATATGGCGCTCAATGGTTTAAAGAGTAGTCTCCCAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTGCTGGTGTGATGCGAAGTCAACAGCAAGCTCTTGGCACAGCTATTAT 300
|||||

```

```

Db      81 AlaAsnAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CGTGAAGCTGTAAGAACTTTAGCAGCTGTCGTAATCCCTATGCGCATTAACCTGGCATA 360
Db 101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTGTTGCTGCTTACTAAAGAACTAAGCGACATTACAAAGCCCTACTCGTGAC 420
Db 121 AspLysAlaValValAlaAlaValGluLeuLysAlaLeuSerValProCysSerAsp 140
QY 421 CAAAAAGAAATAGCTCAAGTTGGACCACTTTCTGCAAACTCTGATACAAATAGGTAAT 480
Db 141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160
QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAGAGGAGTGTATACAGTTGAGGAAGCT 540
Db 161 LeuIleAlaGlnAlaMetAspLysValGlyGluGlyValIleThrValGluAspGly 180
QY 541 AAAGGCTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACTC 600
Db 181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATCTTTGTAATACTATCTCGAAGAAATGGTTTGTGAACCTGATACCACTTATATC 660
Db 201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheile 220
QY 661 CTTTGTAAATCAGAAAGAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGT 720
Db 221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACCTAGAGGTGAAGCACTTGCA 780
Db 241 AlaLysAlaGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATCTGTAGTAAATAGCTCCGTTGAGCACTCCAGTTGTAGCCGTAAAGCTCCTGGT 840
Db 261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280
QY 841 TTTGGTGAAGCGCTGAAGCTATGCTTGAAGATATCTATCTCTTACTCTGAGGAGGAAGCA 900
Db 281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300
QY 901 ATATTTCAAGATCTGCTGTATATAAAGCTTGAAGCTTGTCTCTTCTTTAGGAACAGCT 960
Db 301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320
QY 961 AAAGCTGTATGTATGAAGAAATPACTATCTATCTGCTGATGCTGCTGGAATACGAA 1020
Db 321 LysArgValValIleAsnLysAspThrThrThrIleAspGlyValGlyAspGluAla 340
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAAAGCTCAGATTAT 1080
Db 341 AlaIleGlnGlyArgValThrGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360
QY 1081 GATCGTGAAGAACTTCAAGAACTGCTTTCGAAACCTTGTGTGGAGTAGCTGTTATCCAT 1140
Db 361 AspArgGluLysLeuGlnIleArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTCGAGCTGCTACTGAAACTGAATGAAGAGAGAGAGGATCGTGTAGAGAGTCTCTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysLysLysAlaArgValGluAspAlaLeu 400
QY 1201 AATCCAAAGAGCTCGGTTGAAGAGGATTTCTCCCTGCTGCTGCTGCTGCTTTTGTG 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420
QY 1261 CGCTCCATTAAAGCTCTTGATGATATTAAACCTCTGATGATGATGATGATGCTGCTGCT 1320
Db 421 ArgValAlaSerLysIleAlaAspLysLysGlyGlnAsnGluAspGlnAsnValGlyIle 440
QY 1321 AATATCATCGCTGCTTCTTTGAAGAGCTTTTACGTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db 441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

```

Qy	1381	GAGGTTCTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGGTTTGTGAATTAATCGT	1440
Db	461	GluproSerValValAlaAAsnThrValValAlaGlyAspGlyAsnThrGlyTyrAsnAla	480
Qy	1441	GCATCAGGAGAATAATGAAGACCTTATTAAAGCTGGTGTCAATTCATCTCTAAAAAGATTACA	1500
Db	481	AlaThrGluGluTyrGlyAsnMetCilAspMetGlyIleLeuAspProThrLysValThr	500
Qy	1501	CGTATTTCGATTACAAAATGCGACATCATGATAGCTTCCTTACTTCTAACTACAGAAATCGCGT	1560
Db	501	ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet	520
Qy	1561	ATTCTCGAAAAACCAGAACCTTAAAAAGATATGCTTATGCTCGC-----GGTGGTATG	1614
Db	521	ValThrAspLeu-----ProLysSerAspAlaProAspLeuGlyAlaAlaGlyGlyMet	538
Qy	1615	GGTGGTATGGGTGGTATG	1632
Db	539	GlyGlyMetGlyGlyMet	544

RESULT 13

RESONI TS
AAB19080

ID AAB19080 standard; protein: 545 AA.

XX
XX

AC AAB19080;

XX

DT 08-FEB-2001 (first entry)

XX

DE Amino acid sequence of a 60 kDa protein from *Campylobacter jejuni*.

XX

KW Antigenic protein; fl

8	XX	8
7	XX	7
6	XX	6
5	XX	5
4	XX	4
3	XX	3
2	XX	2
1	XX	1

OS Campylobacter

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

PN EP1043029-A1.

XX
11 - OCT 2000

PD 11-OCT-2000.
yy

XX
PF
03-APR-2000: 2000EP-

03-APR-2000; 2000EP-0201203.
PF
XX

XX
PR 09-APR-1999- 99EP-0201086

FR 09-APK-1999; 99EP-0201086.
XX

PA (ALKU) AKZO NOBEL NV.

XX
XX
XX / ALICE NOBEL NV,
XX

PI Jacobs AAC, Van Den Bosch JF, Nuijten PJM:

[illegible]

DR WPI; 2000-589147/56.

XX

XX

PT Novel *Campylobacter* antigenic proteins, useful for the production of

PT vaccines for protection

XX

PS Disclosure; Page 12-14; 21pp; English.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

The present sequence represents a 60 kDa protein

of C. jejuni protein after incubation of the bacteria in the jejunum. The protein is antigenic, and is visible

of *C. jejuni* protein after incubation of the bl against a flagellin mutant of *C. jejuni* but

against a flagellaless mutant of *C. jejuni* but not incubation with antibodies against wild-type *C. jejuni*.

Incubation with antibodies against wild

QY	7	TCTAAAGAAATCCTTTTGTAGCTAAAGCCCGTGAAGAACTTTTACAGAGGTGTAGATAAA	66
Db	2	AlAlysGluIleIlePheSerAspGluAlaArgAsnLysLeuTyrGluGlyValLysLys	21
QY	67	CTTCGAATGCTGTTAAAGTHAACACTTGGACCTAAAGCCGTATGTCTGTTATTGAAAG	126
Db	22	LeuAsnAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys	41
QY	127	TCTTTTGGTCTCCACAGTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTTGAACTT	186
Db	42	SerPheGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluValGluLeu	61
QY	187	GAGATAAAGTTTGAATAATATGGCGCTCAATAGTGTAAAGAGTAGCTCCCAAACTAGC	246
Db	62	LysAspSerLeuGluAsnMetGlyAlaSerLeuValArgGluValAlaSerLysThrAla	81
QY	247	GATATTCTGCTGATGAACTACAAACAGCAACAGCTCTGCACAAAGCTATTATTCGTGAA	306
Db	82	AspGlnAlaGlyAspGlyThrThrAlaThrValLeuAlaHisAlaIlePheLysGlu	101
QY	307	GGTGTAACCTTGTAGCAGCTGGTGTAAATCTTATGGCCATTAACCTGGCATAGATAAA	366
Db	102	GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys	121
QY	367	GCTGTTGCTGTTACTAAAGAACATAAGCAACATTAACAAGCCTACTCTGTGACCAAAA	426
Db	122	AlaCysGluAlaIleValAlaGluLeuLysLeuSerArgGluValLysAspLysLys	141
QY	427	GAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAACTAGGTAAATATCAT	486
Db	142	GluIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluLysIleGlyAsnLeuIle	161
QY	487	GCTGAAGCTATGGCTAAAGTTGGAAGAGGTGTATTATCAGTTGAGGAAAGCTAAAGGT	546
Db	162	AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluProLysSer	181
QY	547	CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTCACCGTGCTACCTCTCCCA	606
Db	182	IleAsnAspGluLeuAsnValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro	201
QY	607	TACTTTGTAACATACTCTGAGAAAAGTTTGTGAACCTTGATACACCTTATATCTTTGT	666
Db	202	TyrPheIleThrAsnAlaGluLysMetThrValGluLeuSerSerProTyrIleLeuLeu	221
QY	667	AATCAGAAAAGATTACTACATCAAGACATGCTACCAATCTTAGACCAAGTTGCTAAA	726
Db	222	PheAspLysLysIleThrAsnLeuLysAspLeuLeuProValIleGluGlnIleGlnLys	241
QY	727	GTAACCGCTCCACTCCTTATTATTGCTGAAGACGTAGAAGTGAAGCACTTGCAACACTT	786
Db	242	ThrGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu	261
QY	787	GTAGTCAATGAAGCTCCGTGAGACACTCCAAGTTGTAGCCGTAAAGCTCTGTGTTTGGT	846
Db	262	ValValAsnLysLeuArgGlyValLeuAsnIleSerAlaValLysAlaProGlyPheGly	281
QY	847	GAACGCCGTAAAGCTATGCTTGAAGATATTGCTTATCTTACTCGAGAGAGCAATATTT	906
Db	282	AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluValIleSer	301
QY	907	GAAGATCGGTATAAAGCTTTGAAATGTGAAGCTTTGCTTCTTCTTTAGGAACAGCTAA	966
Db	302	GluGluLeuGlyArgThrLeuGluSerAlaThrIleGlnAspLeuGlyGlnAlaSerSer	321
QY	967	GTAGTTATTGACAAAAGAAATATCTATCTGTTGATGTGCTGCTGCAAAATCAGAAGAT	1026
Db	322	ValIleIleAspLysAspAsnThrThrIleValAsnGlyAlaGlyGluLysAlaAsnIle	341
QY	1027	AAAGCTCGAGTTAAACAATTCGTGCAAAATTTGAAGAAACAAGCTCAGATTATGATCGT	1086
Db	342	AspAlaArgValAsnGlnIleLysAlaGlnIleAlaGluThrThrSerAspTyrAspArg	361

201	SerProTyrPheIleAsnLysProGluThrClyAlaValGluLeuGluSerProPheIle	220
661	CTTTGTAATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTTAGAACAAAGTT	720
221	LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal	240
721	GCTAAAGTAAACCGTCCACTCTTATTATTGCTGACGACGTAGAGGTGAAGCACTTGCA	780
241	AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla	260
781	ACACTGTGAGTCAATAAGTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCCCTGGT	840
261	ThrLeuValValAsnThrMetArgLysValAlaValLysAlaProGly	280
841	TTTCGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCCTTACTGGAGGAGACCA	900
281	PheGlyAspArgAsGlyLysAlaMetLeuGluAspIleAlaThrLeuThrGlyGlyThrVal	300
901	ATATTTGAAGATCGTGGTATAAAGCTTCAAAATGTAAAGCTTGCTCTTTAGGAACAGCT	960
301	IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluLeuLeuGlyGluAla	320
961	AAACGCTGATGTTATTGACAAAGAAAATACTACTTCGTGATGCTGTGGAATAACGAA	1020
321	LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla	340
1021	GATATTAAAGCTCGAGTTAAACAAATTCGTGCCAAATTTGAAGAAACAAGCTCAGATTAT	1080
341	AlaIleIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr	360
1081	GATCGTGAAAAACTTCAAGAACGCTTTCGAAAACCTGCTGTGTGAGTAGCTGTTATCCAT	1140
361	AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys	380
1141	GTTGGAGCTGCTACTGAAACTCAATGAAGAGAGAGAGATCGTGTAGAGATGCTCTA	1200
381	ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaAspValGluAspAlaLeu	400
1201	AATCCACACAGAGCTCGGTTGAAGAGGTATTGCTCCCTGGTGGTGTACTGCTTTGTC	1260
401	HisAlaThrArgAlaAlaValGluGluGlyValAlaGlyGlyGlyValAlaLeuIle	420
1261	CGCTCCCATTTAAAGTCCTGTGATGATTAAACCTGCTGATGATGAACCTTGTCGACTT	1320
421	ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle	440
1321	AATATCATCGTCTCTCTTGAAGAGCTTTAGTCAAAATTGCTGCAAAATGCTGGCTAT	1380
441	LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu	460
1381	GAAGGTCTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGTTTTGGATTAAATGCT	1440
461	GluProSerValValAlaAsnThrValLysGlyAspGlyAsnIleGlyTyrAsnAla	480
1441	GCATCAGGAGAAATGAAGACCTTTATTAAGCTGGTGTCAATTGATCTCTCAAAAAGTTACA	1500
481	AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr	500
1501	CGTATTCGATACAAATGCAGCATCAGTAGCTCCCTTACTTCTTAACCTACAGATGGCT	1560
501	ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet	520
1561	ATTGCTGAAAAACACAGAACCTTAAAAAGATATGCTATGCTGCTGGC-----GGTGGTATG	1614
521	ValThrAspLeu-----ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet	538
1615	GGTGGTATGGTGTATGACGGTATG	1641
539	GlyGlyMetGlyMetGlyMetGlyMet	547

[illegible]

Db 81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleTle 100
QY 301 COTGAAGGTGTAACCTTGTAGAGCTGCTGTATCCCTATGCGCATTAACAGTGGCATA 360
Db 101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTGTGTCTGCTTACTTAAGAAGCAATGAGCAGATACAAAGCTCTCTGAG 420
Db 121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140
QY 421 CAAAGAAGAAATAGCTCAAGTGTGAACCACTTCTGCAAACTCTGATACAAATAGGTAAT 480
Db 141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160
QY 481 ATCATAGCTGAAGCTAGCTAAAGTGGAAAGAGGTGTATACAGTTGAGGAGACT 540
Db 161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
QY 541 AAAGGTCTTGAAACTACATAGATGTGTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
Db 181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyThrLeu 200
QY 601 TCTCCATCTTTGTAACTAATCTGAGAAATGGTTTGTGAAGTGTGAAGTGTGAAGTGTGA 660
Db 201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220
QY 661 CTTTGTAAATGAGAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
Db 221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240
QY 721 GCTAAAGTAAACCGCTCCACTCTTATTATTGCTGAAGACGTAGAGTGAAGTGAAGTGA 780
Db 241 AlaLysAlaGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACACCTGTAGTCAATGAAGCTCCGTGGAGCACTCCAAAGTTGTAGCCGTAAAGCTCCTGT 840
Db 261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280
QY 841 TTGTGTGACCGGTAAAGCTATGCTTGAAGATTTGCTATCTTACTGAGGAGAGCA 900
Db 281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyThrVal 300
QY 901 ATATTTGAAGATCGTGTGTAAGCTTGAAGTGAAGTGTGCTTCTTTAGGAACAGCT 960
Db 301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320
QY 961 AAACGTGTAGTTATGACAAAGAAATACTACTACTGCTGATGCTGCTGCTGCTGCTGCTG 1020
Db 321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluAla 340
QY 1021 GATATTAAAGCTGAGTTAAACAAATTCGTGACAAATTCGAGAAACAAAGCTCAGATTAT 1080
Db 341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360
QY 1081 GATCGTCAGAACTCTCAAGACGCTCTGCAAACTGCTGTGAGTGTGCTGCTGCTGCTGCT 1140
Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGGATCGTGTAGAGATGCTCTTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
QY 1201 AATGCAACAAGAGCTGGTGAAGAGGATTTCTCCCTGGTGTGCTGCTGCTGCTGCTGCTG 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyValAlaLeuIle 420
QY 1261 CGCTCCATTAAAGTCTGTGATTAACCTGCTGATGATGATGATGATGATGATGATGATG 1320
Db 421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440
QY 1321 AATATCATCGCTCTCTTGAAGAGCTTCTAGCTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db 441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

QY 1381 GAAGGTCTTATTGTTGTAGAAAAAGTTCTGTGAACCAAAAGATGTTTGGATTTAATGCT 1440
Db 461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGGAGAAATATGAAGACCTTATAAAGCTGTGTGCTCATTCCTAAAAAGTTACA 1500
Db 481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500
QY 1501 CGTATGCTATTCAAAATGCAGCATCAGTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
Db 501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520
QY 1561 ATTGCTGAAAAACCAAGACCTTAAAAAGATATGCTTATGCTGGC-----GGTGGTATG 1614
Db 521 ValThrAspLeu-----ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538
QY 1615 GGTGTATGGTGGTATGACCGTATG 1641
Db 539 GlyGlyMetGlyGlyMetGlyMet 547

Search completed: January 28, 2004, 13:25:21
Job time : 79 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 12:47:27 ; Search time 2467 Seconds
(without alignments)
16225.972 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 1647

Sequence: 1 atggcttctaaagaatcct.....gtatggacggtatgtactag 1647

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST*

1: em_estba:*
2: em_estba:*
3: em_estin:*
4: em_estin:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	571.8	34.7	2265	11	AY109623
2	514.4	31.2	2409	11	BC047350
3	493.6	30.0	2246	11	AK088844
4	469.2	28.5	2290	11	AY104969

RESULT 1
AY109623
LOCUS
DEFINITION
Zea mays CL2221_1 mRNA sequence.
ACCESSION
AY109623
VERSION
AY109623.1
KEYWORDS
HTC.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 2265)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
TITLE
Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL
Unpublished (2002)
REFERENCE
2 (bases 1 to 2265)
Coe, E.H.
TITLE
Direct Submission
JOURNAL
Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT
If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

ALIGNMENTS

5	460.2	27.9	2338	11	AY103778	AY103778 Zea mays
6	394.4	23.9	1376	11	AY106465	AY106465 Zea mays
7	394.4	23.9	2143	11	AY108560	AY108560 Zea mays
8	350.4	21.3	1201	9	AL559862	AL559862 AL559862
9	323.4	19.6	1090	12	BM799922	BM799922 AGENCOURT
10	321	19.5	1327	28	BM770613	BM770613 LLMGTAG37
11	320.6	19.5	1387	11	AY108518	AY108518 Zea mays
12	318.8	19.4	1120	9	AL515262	AL515262 AL515262
13	314.8	19.1	1201	9	AL518632	AL518632 AL518632
14	312.6	19.0	1201	13	EX439563	EX439563 BX439563
15	312.2	19.0	878	10	BF275584	BF275584 GA EB002
16	311.2	18.9	810	14	CD445821	CD445821 ELQ1T0203
17	311	18.9	1201	9	AL557150	AL557150 AL557150
18	310.4	18.8	1201	9	AL515579	AL515579 AL515579
19	308.8	18.7	1061	9	AL532233	AL532233 AL532233
20	305.6	18.6	1201	13	EX446493	EX446493 BX446493
21	304.8	18.5	1201	13	EX462128	EX462128 BX462128
22	303.4	18.4	1201	9	AL557181	AL557181 AL557181
23	299	18.2	1119	12	BM458421	BM458421 AGENCOURT
24	297.6	18.1	746	13	BU003474	BU003474 QG35F17.
25	296.6	18.0	815	14	CB976820	CB976820 CAB40003
26	296	18.0	1201	9	AL514684	AL514684 AL514684
27	295.4	17.9	1051	12	BM478208	BM478208 AGENCOURT
28	293.2	17.8	946	13	BQ882339	BQ882339 AGENCOURT
29	288.4	17.5	1136	13	BU507898	BU507898 AGENCOURT
30	288.2	17.5	1016	12	BM450130	BM450130 AGENCOURT
31	287.8	17.5	804	14	CD434016	CD434016 ELQ1N0317
32	286.8	17.4	790	10	EG599254	EG599254 EST504154
33	286.6	17.4	1201	13	EX460128	EX460128 BX460128
34	286.2	17.4	1086	13	EX404534	EX404534 BX404534
35	286	17.4	876	13	BQ231270	BQ231270 AGENCOURT
36	285.2	17.3	1016	12	BM472194	BM472194 AGENCOURT
37	284.8	17.3	977	10	EG321293	EG321293 Zm04_0490
38	284.4	17.3	1055	12	BM471035	BM471035 AGENCOURT
39	283.4	17.2	895	13	BQ216655	BQ216655 AGENCOURT
40	281.6	17.1	811	14	CB656065	CB656065 OSJNEC09P
41	280.8	17.0	864	14	CB634737	CB634737 OST1EB14H
42	280.4	17.0	850	14	CA792748	CA792748 AGENCOURT
43	279.8	17.0	744	12	BI176450	BI176450 EST517395
44	279.8	17.0	1201	13	EX421838	EX421838 BX421838
45	279.8	17.0	1201	13	EX425282	EX425282 BX425282

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schmale, Iowa State, then clones may be requested from ZmDB:

www.zmdb.iastate.edu.
Location/Qualifiers

FEATURES

source

1. .2265

/organism="Zea mays"

/mol_type="mRNA"

/db_xref="MaizeDB:630867"

/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 615 a 430 c 599 g 538 t 83 others

ORIGIN

Query Match 34.7%; Score 571.8; DB 11; Length 2265;

Best Local Similarity 59.3%; Pred. No. 2.3e-123;

Matches 964; Conservative 0; Mismatches 659; Indels 3; Gaps 1;

```

QY      4  GCTTCTAAAGAAATCCTTTTTCATGCTAAAGCCCGTGAAGAACTTTTCAGAGGTGTAGAT 63
DB      290 GCTGCGAAGGACATCAAGTTTGGTGTGAGGCCGCTGCTCTGATGNNNNNNNNNNNN 349

QY      64  AAACCTTGCAGAACTGTTTAAAGTACACCTTGGACCTTAAAGCGCGTAATGCTGTTATTTGAA 123
DB      350 NNNNNNNCAGATGCTGTCAAGTGACAAATGGACCTTAAAGGCGCAATGTGGTTATTGAG 409

QY      124 AAGCTTTTGGTTCGCCAGTATATACAAAGATGTTGTTATCTGTTCCAAAGAAATTTGAA 183
DB      410 CAAAGCTTTGGTGCAACCGCAAGTCAAAAGAGTGTGTGACTGTAGCAAGAGCAATTGAA 469

QY      184 CTTGAGATAAGTTTGAAATATGGCGCTCAAAATGTTTAAAGAGTAGCTCCCAAACT 243
DB      470 TTTAGGATAGATTAAGAAATGTTGTTGCAAGCTTGTGAACAGTTGCTATGCAACT 529

QY      244 AGCGATATTGCTGGTGAGGAACTACACAGCAACAGTCTCTTGCAACAGCTATTATTCGT 303
DB      530 AATGACACTGCTGGCGATGGTACCACATGTGCCACTGTTTGCACAAAGCAATATTACT 589

QY      304 GAAGTGTAAACTTGTAGCACTGCTGTTATCTTATGCCATTAAGCTGCAATAGAT 363
DB      590 GAGGGGTGCAAACTCTGTGCGCTGGAATGAATGCAATGGAATTAAGGCGTGAATCTCA 649

QY      364 AAAGCTGTTGTTGCTGTACTTAAAGAACTAAGCGCAATTAAGAGCTTACTCTGTGACCAA 423
DB      650 ATGGCTGTTGACGCTGTGTGACCAATCTCAAGGCGATGGCCAGATGATCAGCACTTCA 709

QY      424 AAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAAATAGTAAATATC 483
DB      710 GAAGAAATTTGCAAGGTGGGTACAAATATCAGCAATGGGCAAGGGAATTTGGTGAGCTT 769

QY      484 ATAGCTGAAGCTATGCTTAAAGTTGCAAGAGGAGTGTATCACACTTGGAGAGCTAAA 543
DB      770 ATTGCCAAGGCTATGAGAGAGGTTGGCAAGAGGGTGTATCATCTATTGCGGACGGTAAC 829

QY      544 GGTCTTGAACACTACATTAGATGTGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCT 603
DB      830 ACCCTTTTAAATGAGCTTGAAGTTTGGAGGGTATGAAGTATGACAGAGGTTATCATCTCT 889

QY      604 CCATCTTTGTAACTAATCTCTGAGAAATGGTTTGTGAACCTTGATTAACCTTTATCTCT 663
DB      890 CCGTACTTCTATTAACCAACTCAAAAGCCCAAGAAATGTGAATTTGAAGACCCCAATGATCTTA 949

QY      664 TGTAATCAGAAAAGATTTACTACATGAAAGACATGCTACCAATCTTAGAACAAGTTGCT 723
DB      950 ATTCATCAGAAAGGTTAGCAATATGCAATGCTGTGGTTAAGTTTTAGAAATGGCTCTG 1009

QY      724 AAAGTAAACCGTCCACTCTCTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACA 783

```

```

DB      1010 AAGAAACAAAGCCCTCTACTATTGTTGAGAGAGTGTGGAAGTGAAGCAATTTGGCACT 1069
QY      784 CTTGTAGTCAATTAAGCTCGTGGAGCACTCCAAAGTTGTAGCGTAAAGCTCTCTGTTTT 843
DB      1070 TTGATTATTAACAAGCTTCGTGAGGCACTCAAGGCTGTGTGCTGTCAAGCTCTCTGTTTT 1129
QY      844 GGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCTCTTACTGGAGAGCAATA 903
DB      1130 GGGGAAACAGGAGGCAAACTTACAGACCTTGAACCTTCTTACTGGAGAGGTAATA 1189
QY      904 TTTGAAGATCTGTGTATTAAGCTTGAAGCTTGAAGCTTGTCTTCTTTAGGAACAGCTAAA 963
DB      1190 ACTGAAGAACTAGGAATGAACCTTGAAGCTTGTGAGCTTCAATGCTGGGTATCGNNN 1249
QY      964 COTGTAGTTATTGCAAAAGAAATCTACTATCTGTTGAGTGTGCTGGAAGTCAAGAT 1023
DB      1250 NNGGTGACTGTCTCTAAGGATGACACTGTATTCTTTGATGGAGCCGAGCAAGAGTCC 1309
QY      1024 ATTAAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGAT 1083
DB      1310 ATTGAAGAGAGGCGAGAGCAATTTAGATCAGCAATTTGAAATAGCACTTCAAGATTATGAT 1369
QY      1084 CGTGAAGAACTTCAAGAACTCTTGCACAACTTTGTTGGTGGAGTAGCTGTTATCCATGTT 1143
DB      1370 AAGGAAAGCTCCAGAGCGGTTGGCAAGCTCTCTGGAGGTGTGCTGTTCTGAAGATT 1429
QY      1144 GGAGCTGCTACTGAAACTGAAATGAAGAGAGAGAGGATCGGTGAGAGATCTCTAAAT 1203
DB      1430 GGAGAGCGCCAGCAAGCAAGTGTGTGAGAAAGAGGATAGAGTGAAGTCAAGTCACTGAAT 1489
QY      1204 GCAACAAAGAGCTGCGTGTGAAGAGGATTTGTCCTGCTGGTGGTACTCTTTTGTCCGC 1263
DB      1490 GCTACTAAAGCTGCTGTTGAAGAGGATTTGTTACAGGTGGTGGTGTCTCTCTAT 1549
QY      1264 TCCATTAAAGTCTTGTGATGATTAATAAAGCTGCTGATGATGATGAACTTGTGAGACTTAAT 1323
DB      1550 GCATCGAAGGAGCTTGATAAATTCAGACAGAGCACTTCGATCAGAAAGATTGGTGTGCAA 1609
QY      1324 ATCATCCGTCGTTCTTGAAGAGCTTTTACGTCAAAATGCTGCAATGCTGCAATGCTATGAA 1383
DB      1610 ATCATTCAGAAATGCTTGAAGACACCTGTACACAAATGCTCAATGCTGAGGTGGAG 1669
QY      1384 GGTCTTATTGTTGTAGAAAAGTTGTTGAACCAAAAGA---TGGTTTGGATTAAATGCT 1440
DB      1670 GGAGCAGTAGTTGTAGAAAGCTTTTGGAGCAAGAAATACTGACCTGGGTATGATGCT 1729
QY      1441 GCATCAGGAGATATGAAGACCTTATTAAGCTGTTGTTGATGATGATGATGATGATGATGATGAT 1500
DB      1730 GCTAAAGGTGAATATTTGACATGTTGAAGACCGGTATCATTTGACCCGCTAAAGGTTATC 1789
QY      1501 CGTATTGCAATTAACAAATGCAGATCAGTACCTCTCTTCTTCTTAACTACAGAAATCGCT 1560
DB      1790 AGAAGCTGTTGGTGGATGCTGCTAGTGTGCTGCTCTCTGATGACAAACCGGAAATCAATA 1849
QY      1561 ATTGCTGAAAAACCAAGAACTTAAAGAGATATGCTTATGCTGCGCGGTGATGAGGTGCT 1620
DB      1850 ATTGTAGATTTCCCAAGGAGGACCGGCTCCAGCAATGGTGGCGCATGGGTGGC 1909
QY      1621 ATGGGT 1626
DB      1910 ATGGAT 1915

```

RESULT 2

BC047350

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC047350 2409 bp mRNA linear HTC 28-FEB-2003
Homo sapiens, clone IMAGE:4214709, mRNA.

BC047350

BC047350.1 GI:28611161

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 2409)
Straussberg, R.
Direct Submission
Submitted (28-FEB-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: c9apbs-remail.nih.gov

Tissue Procurement: David N. Louis, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 26 Row: h Column: 17
This clone has the following problem: no 5' EST match.

FEATURES
source

1..2409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4214709"
/tissue_type="Brain, anaplastic oligodendroglioma with
1p/19q loss"
/clone_lib="NCI CGAP_Brn67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
774 a 420 c 559 g 656 t

BASE COUNT
ORIGIN

Query Match 31.2%; Score 514.4; DB 11; Length 2409;
Best Local Similarity 57.5%; Pred. No. 7.5e-110;
Matches 945; Conservative 0; Mismatches 696; Indels 3; Gaps 1;
6 TTCTAAAGAAATCCTTTTGTGATGCTAAAGCCCGTGAAACCTTTCACAGGTGTAGATAA 65
285 TGCCAAAGATGTAATAATTTGGTGCAGATGCCCGAGCCTTAATGCTTCAAGGTGTAGACCT 344
66 ACTTGCATAATCCTGTTAAAGTAACACTTGGACCTAAAGCCGTAATGTCGTTATTGAAA 125
345 TTTAGCCGATGCTGTGGCCGTTCATAGGGCCAAAGGGAAGACAGTGATTTAGACA 404
126 GTCTTTTGGTCCCAAGTTAACAAGAGTGTGTATCTGTTGCAAAAGAAATTTGAAC 185
405 GAGTTGGGGAAGTCCCAAGTAACAAGAGTGTGTGACTGTTCGAAAGTCAATGTACT 464
186 TGAAGATAAGTTGAAATATAGGCGCTCAATGTTAAAGAGTAGTCCCAAAATAG 245
465 AAGAAGATAATACAGAACAATTGAGAGTAAACTTTGTTCAAGATGTGCAATAACACAAA 524
246 CGATATTGCTGTGATGGAACCTCAACAGCAACAGTCTTGCACAGCTATTTATCGTGA 305
525 TGAAGAAGCTGGGATGGCACTACCTGCTACTGTGACGACGCTCTATAGCCAAGGA 584
306 AGGTGTAATACTTTAGCAGCTGCTGTAATCTTATGCCATTAAGAGCTGGCATAGATA 365
585 AGGCTTCAGAGAAGTTAGCAAGGTGCTAATCCAGTGAATCAGGAGAGGTGTGATGTT 644
366 AGCTGTTGCTGTTACTAAGAACTAAGGCAATTACAAAGCTACTCGTGACCAAAA 425
645 ATCTGTTGCTGCTGTAATTTGCTGAACTTAAAGCAGTCTAAACCTGTGACCCACCCCTGA 704

QY 426 AGAATAGCTCAAGTTGNAACCACTTCTGCAAACTCTGATACAACTATAGGTAAATCAT 485
DB 705 AGAAATTCACAGGTTGCTACATTTCTGCAAAACGAGGACAAAGAAATTTGCAATATCAT 764
QY 486 AGCTCAAGCTATGGCTAAAAGTTGGAAGAGGAGTGTATACAGTTGAGGAGCTAAAGG 545
DB 765 CTCGATGCAATGAAAGAAAGTTGGAAGAGGAGTGTATACAGTTAAGAGATGGAAGAAC 824
QY 546 TCTTGAAACTACATTTAGATGTTGTTGAAGGAATGAAGTTGACCGTGGCTACTCTCTCC 605
DB 825 ACTGAATGATGAATTTAGAAATTTATTGAAGGCATGAAGTTTGTATCGAGGCTATTTCTCC 884
QY 606 ATACTTTCTAATCTGAGAAATGGTTTGTGAACCTGTATACCCCTTATATCTTTG 665
DB 885 ATACTTTTATATACATCAAAAGGTCAGAAATGTGAATTTCCAGGATGCTATGTTCTGT 944
QY 666 TAATGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTTAGAACAAAGTTGCTAA 725
DB 945 GAGTGAAAGAAATTTCTAGTATCCAGTCCATTTGCTCTTGAATTTGCCAATGC 1004
QY 726 AGTAAACCGTCCACTCTTATTATTGCTGAAGACGTGAAGGTGAAGCACTTTGCAACACT 785
DB 1005 TCACCGTAAAGCCTTTGGTCAATAATCGCTGAAGATGTTGATGGAGAAGCTCTAAGTACACT 1064
QY 786 TGTAGTCAATAGCTCCGTTGAGCACTCCAAGTTGTAGCCGTAAAGCTCTCTGTTTGG 845
DB 1065 CGTCTTGAATAGGCTAAAGGTTGCTTTCAGGTTTGTGCGAGCTCCAGGCTTCCAGGTTTGG 1124
QY 846 TCAACCGCTAAAGCTATGCTTTGAAGATTTGCTATCTTACTGAGGAGAGCAATATT 905
DB 1125 TGACATAGAAAGAACCACTTAAGATATGCTATTTGCTACTGTTGTTGCTGCTTTGG 1184
QY 906 TGAAGATCGTGA---TAAAGCTTGAAGTGTAAAGCTTGTCTCTTTAGGAACAGCTAA 962
DB 1185 AGAAGAGGATTTGACCCCTGAACTTTGAAGACGTTTCAAGCTCATGACTTTAGGAAAGTTGG 1244
QY 963 ACCTGTAGTTATTGACAAAGAAATACTACTATCGTTGATGCTGTGAAATCAGAGA 1022
DB 1245 AGAGTCAATTGTGACCAAGACGATGCCATGCTCTTAAAGGAAAGGTTGCAAGGCTCA 1304
QY 1023 TATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCAGAAACAAAGCTCAGATTATGA 1082
DB 1305 AATTGAAAGCTATTCAAGAAATCATTTGACAGTTAGATGTCACTACTAGTAATATGA 1364
QY 1083 TCGTGAAGAACTTCAAGACGCTTTGCAAACTTTGTTGGTGAAGTGTGTTATCCATGT 1142
DB 1365 AAAGGAAAACTGAATGAACGCTTTGCAAACTTTTCAAGTGAAGTGTGCTGCTGAGGT 1424
QY 1143 TGGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGATGCTGTAGAGATGCTCTTAAA 1202
DB 1425 TGGTGGGACAGTGATGTTGAAGTGAATGAAAGAAAGACAGAGTTACAGATGCCCTTAA 1484
QY 1203 TGCAACAGAGCTCGGCTTGAAGAGGTTATGTCCTGCTGCTGCTGCTGCTTTTGTCCG 1262
DB 1485 TGCTACAGAGCTGCTGTTGAAGAGGCACTTTTGGAGGGGGTGTGTCCTCTCTTCG 1544
QY 1263 CTCATTAAAGTCTTGTATGATATTAACCTGCTGATGATGATGAACCTTGTGCACTTAA 1322
DB 1545 ATGATTTCCAGCTTTGACTCATTTGACTCCAGCTTAATGAAGATCAAAATTTGCTATAGA 1604
QY 1323 TATCATCGCTGCTCTTCTGAAGAGCTTTTACGTCAAAATTTGCTGCAAAATGCTGGCTATGA 1382
DB 1605 AATTATTAAAGAACACTCAAAATTTCCAGCAATGACCAATTTGCTAAGAAATGCAAGTGTGA 1664
QY 1383 AGGTTCCTATTGTTAGAAAAGTTGCTGAACCAAAAGATGGTTTTCGATTATATGCTGC 1442
DB 1665 AGGATCTTTGATAGTTGAGAAATTTATGCAAGTTCCTCAGAAAGTTGGTTATGATGCTAT 1724
QY 1443 ATCAGGAGAAATGAGACCTTTATTAAGCTGGTGTCTATTGATCTTCAAAAGGTTTACAG 1502
DB 1725 GGCTGGAGATTTTGTGAATATGTTGGAAGAAAGGAACTATTGACCCCAACAAAGGTTGTGAG 1784

1503 TATTGCATTACAAATGAGCATCAGTAGCCTCTTACTTCTTAACCTACAGAAATGGCTAT 1562
|||||
1785 AACTGCTTTATTGGATGCTGCTGGTGTGGCCCTCTCTGTTAACTACAGCAAGTTGTACT 1844
|||||
1563 TGCTCAAAAACAGAACCTAAAGATATGCTATGCTGCGGTGGTATGGTGGTAT 1622
|||||
1845 CACAGAAATCTTAAGAGAGAGAGACCTGGAATGGTGGTGGTATGGGAGG 1904
|||||
1623 GGGTGGTATGGACGGTATGTTACTA 1646
|||||
1905 TGGTATGGAGGTGGCATCTTCTA 1928
|||||

AK088844 2246 bp mRNA linear HTC 05-DEC-2002
Mus musculus 2 days neonate thymic cells cDNA, RIKEN
full-length enriched library, clone:E430028C20 product:heat shock
protein, 60 kDa, full insert sequence.

AK088844 GI:26353953
HTC; CAP trapper.
Mus musculus (house mouse)

ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

9279253
10349636

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

20493374
11042159

3
Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

20530913
11076961

4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavert, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Stauber, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsi, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gusticich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Washima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohtsuki, S.
and Hayashizaki, Y.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

1. 2246
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM DB:E430028C20"
/db_xref="taxon:10090"
/clone="E430028C20"
/cell_type="thymic cells"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days neonate"
71. 1792
/note="unnamed protein product; heat shock protein, 60 kDa
(MGD|MG1:96242, GI|X53584, evidence: BLASTN, 98%,
match=2111)
putative"
/codon_start=1
/protein_id="BAC40607.1"
/db_xref="GI:26353953"
/translation="MLRLPTVLRQMRVSRALAPHLTRAYAKDVKFGADARALMLOGV
DLADAVATVMGPKRTVITFQSGSPKVTGKVTAKSIDLKDKNKNIAGKLVQDVDA
NTNNEAGDGTATVTLARISAKGFEKISGNAPVEIRGVMLAVDAVIAELKKQSK
FVTTPETIAQTISANGDKDIGNIISDAMKNVGRKGVITVQKLVDAIAHAKPLVIAE
FDVGYISPTINTSKGQCFQDAYVLLISSEKISSVOSIIPALSIANARHPVLIAE
DVGDEALSTVLNLRKVLQVAVKFGDGNRKNQKDMAITAGVAVFGEGEKLNL
EDVADHLGKGVETVTKDAMLLKGGDEAHIEKRIQETITQDITTFSEKELNL
RLAKISGVAVLVKGGTSDVENKPKDRTVDALNATRAAVEEGIVTGGCCALLRCPA
LDSLKPANEDQKIGIEIILKALPTIANTAKNAGVEGSLIVEKILQSSSEVGYDAMLG
DFVNMVKEGILDPKTVRTALLDAAGVASILLTTAEAVVTEIPKEKDPGMGAMGGMG

FEATURES
source

CDS

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2246)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 2246
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM DB:E430028C20"
/db_xref="taxon:10090"
/clone="E430028C20"
/cell_type="thymic cells"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days neonate"
71. 1792
/note="unnamed protein product; heat shock protein, 60 kDa
(MGD|MG1:96242, GI|X53584, evidence: BLASTN, 98%,
match=2111)
putative"
/codon_start=1
/protein_id="BAC40607.1"
/db_xref="GI:26353953"
/translation="MLRLPTVLRQMRVSRALAPHLTRAYAKDVKFGADARALMLOGV
DLADAVATVMGPKRTVITFQSGSPKVTGKVTAKSIDLKDKNKNIAGKLVQDVDA
NTNNEAGDGTATVTLARISAKGFEKISGNAPVEIRGVMLAVDAVIAELKKQSK
FVTTPETIAQTISANGDKDIGNIISDAMKNVGRKGVITVQKLVDAIAHAKPLVIAE
FDVGYISPTINTSKGQCFQDAYVLLISSEKISSVOSIIPALSIANARHPVLIAE
DVGDEALSTVLNLRKVLQVAVKFGDGNRKNQKDMAITAGVAVFGEGEKLNL
EDVADHLGKGVETVTKDAMLLKGGDEAHIEKRIQETITQDITTFSEKELNL
RLAKISGVAVLVKGGTSDVENKPKDRTVDALNATRAAVEEGIVTGGCCALLRCPA
LDSLKPANEDQKIGIEIILKALPTIANTAKNAGVEGSLIVEKILQSSSEVGYDAMLG
DFVNMVKEGILDPKTVRTALLDAAGVASILLTTAEAVVTEIPKEKDPGMGAMGGMG

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

COMMENT

GMGGGMP"
polyA_signal 2225..2230
/note="putative"
polyA_site 2246
/note="putative"
BASE COUNT 671 a 413 c 581 g 581 t
ORIGIN

Query Match 30.0%; Score 493.6; DB 11; Length 2246;
Best Local Similarity 56.7%; Pred. No. 5.7e-105;
Matches 932; Conservative 0; Mismatches 709; Indels 3; Gaps 1;

QY 6 TTCTAAAGAAATCTCTTTTGTATCTAAAGCCGCTGAAAGAACTTTTCAGAGGTCTAGATAA 65
Db 148 TGGCAAGATGTAAATTTGGTGGGACGCTCGAGCCCTTAATGCTTCAAGGTGTACCT 207

QY 66 ACTTGCAAATGCTGTAAAGTAACACTTGGACCTAAAGCCGCTTAATGCTTATTGAAAA 125
Db 208 TTATAGCAGATGCTGTAGCTTTACAATGGGGCCAAAGGAAGAACTGTGATTATTGAACA 267

QY 126 GTCTTTTGGTTCCTCCAGTTATTAACAAGATGTGTATCTGTGGCAAGAAATTAAGACT 185
Db 268 GAGTTGGGAAGTCCCAAGTAACAAGATGGGGTCACTGTGTCAAAGTCAATTTGATTT 327

QY 186 TGAAGATAAGTTTGAAGATATGGCGCTCAATGCTTAAAGAAAGTAGTCTCCCAAACTAG 245
Db 328 AAAGGATAAATAACAATAATTTGAGCTAAACTTTTCAGAGCTTGCCTAATTAACAAA 387

QY 246 CGATATTTGCTGTGTGGAAGTAACACAGCAACAGCTCTTTGCAAGCTTATTTATCGTGA 305
Db 388 CGAAGAGCTGGGGATGGCAACCACTGCTCCACTGTCTGCGACGATCTATTGCCAAGGA 447

QY 306 AGGTGTAAACTTTGTAGACGCTGGTGTGTATCTTATGCGCCATTAAGCTGGGCTATGATA 365
Db 448 GGGCTTTTGAAGATCAGCAAGGGGCTAATCAGTGGAAATCCGGAGAGGTGTGATGTT 507

QY 366 AGCTGTGTGTGCTTACTTAAAGAACTTAAGCACTTAAGCAAGCTACTCGTGACCAAAA 425
Db 508 GGCCTGTGGATGCTGAATTCCTGAACCTTAAGAAACAGCTCTAAGCTGTGACACCCCTGA 567

QY 426 AGAAATAGCTCAAGTTGGAAACCACTTTCTGCAAACTCTGTATCAACAATAGGTATATCAT 485
Db 568 AGAAATGCTCAGTTGCTCAAAATTTCTGCAATGGAGACAAAGACATTTGGGAACATCAT 627

QY 486 AGCTGAAGCTATGCTTAAGTTGAAAGAGAGGTGTATCACAGTGTGAGAAAGCTTAAGG 545
Db 628 TTCTGTATGCAATGAAAGAGTTTGAAGAAAGGTGTCTACACAGTGAAGGATGGAAGAAC 687

QY 546 TCTTGAACCTACATTAGATCTGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
Db 688 CCTGATGATGAGCTAGAAATTTATGAGGCAATGAAGTTTGTATAGAGATATATTCCCC 747

QY 606 ATACTTTGTAACTAATCTCGAGAAATAGTTTGTGAACCTTGATAACCTTATATCTTTG 665
Db 748 GTATTTTATTAACACATCAAAAGGTCAAAAGTGTGAATTCCAAGATGCTTATGCTTTGTT 807

QY 666 TAATGAGAAAAGATTAAGTGTGAAAGAGAGGTGTATCACAGTGTGAGAAAGCTTAAGG 725
Db 808 GAGTGAAGAAAGAAATTTCCAGTGTTCAGTCCATTCCTGCTCTTGAAATTTGCTAATGC 867

QY 726 AGTAAACCGTCCCTCTTATTTATGCTGAAGAGCTAGAGGTGAAGCACTTGAACACT 785
Db 868 TCATCGGAAGCAATGTTGTAATTCGCCGAAGAGCTTGAAGGAGCACTTAAGCAGCT. 927

QY 786 TGTAGTCAATAAGCTCGGTGAGACACTCCAAAGTTGTAGCCGTAAAGCTCTCTGTTTTGG 845
Db 928 GGTTTTGAACAGGCTAAAGTTGCTTCAAGTTGTGCGAGTCAAGCTCCAGGATTTGG 987

QY 846 TGAACGCCGTAAGCTATGCTTGAAGATATGCTTATCTTACTGGAGAGAAAGCAATATT 905
Db 988 GGACAAATAGGAAGAACCAAGCTTAAAGATATGCTTATGCTACTGCTGCTGAGTGTGG 1047

QY 906 TGAAGATCGTGGTA--TAAAGCTTGAAGTGAAGCTTGTCTTTTGAAGAACAGCTAA 962

Db 1048 AGAAGAGGGGTTCAATCTAAATCTTGAAGATGTTCAAGCTCATGACTTAGAAAAAGTTGG 1107
QY 963 AGGTGTAGTATTGACAAAGAAAAATACTACTATCTGTTGATGGTCTGCGAAATCAGAAGA 1022
Db 1108 GGAGGTCAATTGTCAACCAAGATGATGCCATGCTTTTGAAGGAAAAAGGTGACGAAGCTCA 1167

QY 1023 TATTAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGA 1082
Db 1168 CATTTGAAAACGTTTTCAGAAATCACTGAGCAGCTAGACATCACAACTAGTGAATATGA 1227

QY 1083 TCGTGAAGAACTTCAAGAACGCTTTGCAAAACCTTTGTTGGTGGAGTAGCTGTTATCCATGT 1142
Db 1228 AAAAGAAAAGCTGAACGAGCGACTTCTGCTAAACCTTCAGATGGAGTAGCTGTGTTGAAGGT 1287

QY 1143 TGAAGCTGCTACTGAAACCTGAATGAAGAAAGAAAGGATCGTGTAGAGATGCTCTAAA 1202
Db 1288 TGGAGGAACAAGTGTGTTGAAGTGAATGAGAAAAAGACAGAGTTTACTGATGCTCTCAA 1347

QY 1203 TGAACAAGAGCTGCGGTTGAAGAGGTATTGTCCTGTTGGTGGTACTGCTTTTGTCCG 1262
Db 1348 TGCTCAAGAGCAGCTGTTGAAGAGGCACTTGTCTAGAGGGGGCTGCGCTCTGCTTCG 1407

QY 1263 CTCATTAAAGTCTCTTGAATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGAT 1322
Db 1408 GTGCATCCCGAGCTTGGATTCATTAAGCCCTGCTAATGAAGACCAAGAAATAGGTATAGA 1467

QY 1323 TATCATCGGTGCTTCTTGAAGAGCTTTTACGTCAATTTGCTGCAATGCTGCTGCTATGA 1382
Db 1468 AATTATTAAAGAGCACTTAAATTTCTGCAATGACGATTTGCTAAGAAATGACAGGTGTGA 1527

QY 1383 AGCTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGTTTGGATTTAATGCTGC 1442
Db 1528 AGATCTTTGATAGTTGAGAAAAATTCGACAGGTTCTCGAAGTTGTTTATGACGCCAT 1587

QY 1443 ATCAGGAGATATGAAGACCTTATTAAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1502
Db 1588 GCCTGGAGATTTTGTGAACATGTTGGAAGAAAGGATCATTTGATCCAAACAAAGGTTGTGAG 1647

QY 1503 TATTGATTAACAAATGACAGCATCAGTAGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1562
Db 1648 AACTGCTTACTGATGCTGCTGGGGTGGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1707

QY 1563 TGTGTAACCAACAGAACCTTAAAGATATGCTATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCT 1622
Db 1708 GACAGAAATTCCTAAGAGAGAGGACCTGGAATGGGTGCAATGGTGGCATGGGAGG 1767

QY 1623 GGGTGTGATGACGCGTATGACTA 1646
Db 1768 GGGTATGGGAGCGGCATGTTCTA 1791

RESULT 4
AY104969
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE

AY104969
Zea mays
AY104969.1
HTC.
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2290)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2290)
Coe, E.H.
Direct Submission

PC0070942
mRNA
linear
2290 bp
HTC 16-OCT-2002

GI:21208047

JOURNAL

Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES

Location/Qualifiers
1..2290
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:633867"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"

/note="This sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 684 a 457 c 583 g 565 t 1 others

ORIGIN

Query Match 28.5%; Score 469.2; DB 11; Length 2290;
Best Local Similarity 58.0%; Pred. No. 3.1e-99;
Matches 890; Conservative 0; Mismatches 634; Indels 11; Gaps 3;
43 AAACCTTTCACGAGGTGTAGATAAATCTGCAAAATGCTGTTAAAGTAACACTTGGACCTAAA 102
41.4 AAAATGCAAAATGGGTCAACAGCTTGCAGATCTCGTCGGGGTTACTCTTGACCAAG 473
103 GCGCGTAATGCTTATTGAAAGTCTTTTGGTCCCGAGTTATTACAAAAGATGGTGA 162
474 GGGCGAAATGTAGTCTCTGAAAGCAAGTATGGCTCACCTAAATTTGTCAATGATGGTGC 533
163 TCTGTTGCAAAAGAAATGAACCTTGAAGATAAGTTGAAATATGGCGCTCAATGGTT 222
534 ACTGTCGCAAGAGGTGAACTTGAAGACCTGTTGAAATATTTGGTCTAAATTTGGTC 593
223 AAAGAAGTAGCTCCCAAACTAGCGATATTGCTGGTATGGAATCTACCAACAGCAAGTC 282
594 CGCAAGCTGCTGCTAAACAAATGACTTGGCTGGTATGGCAACCACTTCGGTTGTC 653
283 CTTCACAAAGCTATTATCGTGAAGGTGTAACCTTTAGAGCTGGTGGTAAATCCTATG 342
654 CTTCCTCAAGGGATGATTACTGAGGGGTGAAGGTTGTAGCTGCTGGTCTAATCCAGTG 713
343 GCATTAACAGTGGCATAGATAAGCTGTTGCTGTTTACTTAAAGAACTAAGCGACATT 402
714 CAGATACCCGTTGATTGAGAAACAGCCAAAGCACTAGTCAGTGAATCCAAAGATG 773
403 ACAAGCCCTACTCGTGACCAAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCT 462
774 TCCAAGGGAGG - TGAGGACAGTGAAGCTTCTGATGTTGTCAGTGTAGTCGGGAAAT 831
463 GATACAACTAAGTAAATATATAGCTGAAGCTATGGCTAAAGTTGGAAAGAGGAGTGT 522
832 AACTATGAATTTGGGAATATAGTATGAGGCTATGAAAGAGTTGGCGCGCAGGAGTG 891
523 ATCAGATTGAGGAAGCTAAAGCTTCTGAACTACATTAGATGTGTTGAGGAATGAG 582
892 GTCACTTTGAGAGGAAGAGTGTCTGAAATATACCTCTATTCTGTTGANGGAATGAG 951
583 TTGACCGTGGTACTCTCTCCATCTTTGTAATCTTGAATCTTGAAGAAATGGTTTGTGA 642
952 TTGACCGTGGTATATTTCTCCCTACTTTGTGACTGACGTGAAAGAAATGACATCGAG 1011
643 CTTGATAACCTTATCTCTTGTATGAGAAAGATTTAGCATGAAAGACATGCTA 702
1012 TATGAAATTTGCAAGCTTCTCTCTGTCGACAGAAATTAACAATGCCAGATCTTATC 1071

QY 703 CCAATCTTTAGAACAAAGTTGCTAAAGTAACCCCTCCACTCTTATTATTGCTGAAGACGTA 762
DB 1072 ACTATCTCGGAGGATGCTATTAGAAGTGGATATCAATTTCTAATAGTGCAGAGACATT 1131
QY 763 GAAGGTGAAGCACTTTCGAACACTTCTAGTCAATAAGCTCCGTTGGAGCACTCCAAAGTTGA 822
DB 1132 GAGCAGGAAGCTCTTTGCAACCCCTTGTGGTTAATAGGCTTAGAGGTGCATTGAAGATTGCT 1191
QY 823 GCGCTAAAGCTCCTGGTTTGGTGAAACGCGCTAAAGCTATGCTTGAAGATATTGCTATC 882
DB 1192 GCTATTAAGGCCCTGTTTGGAGAGCGCAAAAGTCAATATCTTCATGACATTTGCTACT 1251
QY 883 CTTTACTGAGGAGAGAGCAATATTTCAAGATCTGCTGATATAAAGCTTGAAGATTTGAAGCTTTG 942
DB 1252 CTCCTGAGGACCTGTCTATCAGAGAAGAGTTGGATTATCTCTTAGAAGGACAGCAAT 1311
QY 943 TCTCTTTAGGAACAGCTAAACGCTGTAGTTATTGACAAAGAAATATCTATCTCGTTGAT 1002
DB 1312 GCGGCTCTAGGAACCTGCTTAAGGTTGTAGTTTACCAGGATTTCAACAACTAGTGGGA 1371
QY 1003 GGTGCTGGAATATCAGAGATATTAAAGCTCGAGTTAAACAAATTCGTCGCAAAATTTGAA 1062
DB 1372 GATGCACTTACACAGGAGAGTGAACAAAGGGTGACACAAATCAAGAACCCAGATTGAG 1431
QY 1063 GAAACAAAGCTCAGATTATGATCTGTAAGAACTTCAAGAACGCTTGTGCAAACTTTGTTGT 1122
DB 1432 GCTACTGACAAAGATATGAGAGAGAAATTAACGAGAGATAGCTTAATCTCTGTTG 1491
QY 1123 GGAGTAGCTGTTCATCTGTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGGAT 1182
DB 1492 GGTGTTGCTGTCATTTCAGGTAGGACACAGACAGAACTGAGCTCAAGGAAAGAAATTTG 1551
QY 1183 CGTGTAGAGAGCTCTAAATGCAACAGAGCTCGGTTGGAAGAGGTATTGTCCTGCTGT 1242
DB 1552 AGGTTGAGGATGCCCTGAATGCTACAAAGGCGAGCTCGAGAAAGGTATGCTTGTAGT 1611
QY 1243 GGTGTTGCTGTTTGTGCGCTCCATTAAGTCTTGTGATGATTATAA - - - - - CTTGCT 1296
DB 1612 GGTGCTGTACCTTTTGAGGCTTCATCAAAAGTTGATGCAATTTAGAGACCTTTGAG 1671
QY 1297 GATGATGATGAACCTTGTGAGCTTAAATATCATCCGTTCTCTTTGAAGAGCTTTTACGT 1356
DB 1672 AACGATGAACAGAGAGGTTGGGCTGAAATTTGAAGAAAGTCTCTGAGCTATCCACTAAA 1731
QY 1357 CAAATGCTGCAATGCTGGCTATGAAGTTCTTATTTGTAGAAAAAGTTGCTG - - - - AA 1413
DB 1732 TTGATTGCAAAATGCGAGCGCTTAATGGCAGTGTGTTTACTGAGAAGGTTCTTGCCTAAT 1791
QY 1414 CCAAAAGATGTTTGTGATTTAATGCTCATCAGGAGATATGAAGACCTTTATTAAGCT 1473
DB 1792 GAGAACTTCAGTATGGGTACATGCGGTACAGGGGAGTATGAGGACTTTGATGCTGCT 1851
QY 1474 GGTGCTATGATCTTAAAGAAAGTTACAGTATTTGATTTACAAATTCAGAGCATCAGTAGCC 1533
DB 1852 GGTATCATTTGATCCCAAGGTTGAGGTTGAGGTTGAGGAGCTGCTGATCGTGGCG 1911
QY 1534 TCCTTACTTCTAATCTACAGATGCGCTATTGCTGA 1568
DB 1912 AAAACCTTTCATCACCTCTGATGCCCTGTTGTTGA 1946

RESULT 5

AY103778

LOCUS

DEFINITION Zea mays PC063180 mRNA sequence.

ACCESSION AY103778

VERSION AY103778.1 GI:21206856

KEYWORDS HT.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

2338 bp

mRNA

linear

HTC 16-OCT-2002

clade: Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 2338)
 Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
 Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
 Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 Unpublished (2002)
 2 (bases 1 to 2338)
 Coe,E.H.
 Direct Submission
 Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 If you are interested in getting corresponding physical clones,
 these are publicly available from ZmDB and may be found by BLAST
 searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
 www.zmdb.iastate.edu.

FEATURES
 source
 1..2338
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="MaizeDB:633377"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 668 a 483 c 643 g 544 t

ORIGIN

Query Match 27.9%; Score 460.2; DB 11; Length 2338;
 Best Local Similarity 58.0%; Pred. No. 4.1e-97;
 Matches 893; Conservative 0; Mismatches 633; Indels 13; Gaps 4;

QY 25 GATGCTAAAGCCGCGTGAAGAACTTTCACGAGGTGTAGATAAATTCGAAATGCTGTTAAA 84
 DB 422 GATGGTCACTATCAAGAAGCTCCAGACTGGAGTCAATAAGCTGCAGACCTAGTTGGA 481
 QY 85 GTAACACTTGGACCTAAAGCCGTAATGTCGTTATTGAAAAGTCTTTGGTTCCCGATT 144
 DB 482 GTTACACTCGGACCAAGGGAAGATGTCGTTTTGGAGAGCAAGTACGGCTCTCCCAAG 541
 QY 145 ATTACAAAGATGCTGTATCTGTTGCAAAAGAAATTCGAATGAAGATAAGTTTGAAT 204
 DB 542 ATTGTTAACATGGCGTTACAGTTTGAAGAGAGGTGAGCTGGAGGACCCTGTGGAAC 601
 QY 205 ATGGGCGCTCAAAATGTTAAAGAGTAGCTCCCAAAATAGCGATATGCTGGTGAATGGA 264
 DB 602 ATTGGAGCTAAATTTGGTCAGCAAGCTGCAGTAAAGACCAACGATCTTGTGGAGATGG 661
 QY 265 ACTAACACAGCAAGCTCCCTTGCACAGCTATTTATCGTAAGTGTAAACCTGTAGCA 324
 DB 662 ACAACCACTCTGTGTCCTTGTCTCAGGGGTGATGTCGAGGGGTGTTAAGTTGTGGCA 721
 QY 325 GCTGGTGTATCTATGGCCATTAAACGTGGCATAGATAAGCTGTTGTCTGTACT 384
 DB 722 GCTGGTGTATCTCTGTTAGATTCTCGATTACTCTGTTATCGAGATAACAGCAACTAGT 781
 QY 385 AAGAACTAAGCGACATTACAAAGCTTACTCGTCAACCAAAAGAAATAGCTCAAGTTGGA 444
 DB 782 GAAGAACTACGGAAGTTGCTTAAGAGG---TTGAAGATAGTAGCTTCGAGATGTTGCT 838
 QY 445 ACCATTTCTCAAACTCTGTATACAACTAAGTAAATATCATAGCTGAAGCTATGCGTAAA 504
 DB 839 GCTGTTAGTGTGCGCAACAACTATGAAATTTGTAACATGATAGCAGAGGCCATGAGCAAG 898
 QY 505 GTTGGAAAAGAGGTGTTTACAGTTGAGGAAGCTAAAGTCTTTGAACTACATTAGAT 564

RESULT 6

DB 899 GTCGGTCGGAAGGGTGTGTTTACCTCGAAGAGGGGAGAGTTCTGAGAACCTTCTCTAT 958
 QY 565 GTGGTTGAAGGAATGAAGTTTGACCGTGGCTACTCTCTCCATACCTTTGTAATCTATCT 624
 DB 959 GTTGTGGAGGAATGCAATTTTGACCGTGGTTATCATCTCTCCGTAATTTGTTAAACAG 1018
 QY 625 GAGAAAATGTTTGTGAACCTTGATAACCCCTTATATCTCTTGTAAATGAGAAAAAGATT 684
 DB 1019 GAGAAAATGTTCTCCGAGTATGAGAACTGCAAGCTGCTTTGGTTGCAAAAAGATCACC 1078
 QY 685 AGCATGA-AAGCATGCTACCAATCTTAGAACAAAGTTGTGTAAGTAACCGTCCACTCT 743
 DB 1079 AACGCAAGGGATCTTATCAACCTTTTGGAGAAGCCATCAGAGGTGCATACCAATCTCT 1138
 QY 744 TATTATTGCTGAAGAGCTAGAAGGTGAAGCACTTGCAACACTTTGTAGTCAATTAAGCTCC 803
 DB 1139 GATAATTTGCTGAGGATATTGAGCAGGAGGCTCTTGCTACCCCTTGTCTCAACAGCTCAG 1198
 QY 804 TGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTGGTTTGTGGTGAACCGCGTAAAGCTAT 863
 DB 1199 AGGATCACTGAAGATCGTGCCTCAAGCCCTGTTTGGAGAGGCAAGACTCAGTA 1258
 QY 864 GCTTGAAGATATTGCTATCTCTTACTGGAGAGAGCAATATTTGAAGATCGTGGTATAAA 923
 DB 1259 CTTGGATGATATTGCCATCTTACTGGAGCAACTGTAATCAGAGATGAAGTTGGGCTGC 1318
 QY 924 GCTTGAAGATTAAGCTTCTTCTTTAGAACAGCTAAAGCTGTAGTTATTGACAAAGA 983
 DB 1319 ACTTGAAGAGCTGACAAATTCAGTTCTTGGAAAGCTGCAAGGTTGTCTCTACGAAAGA 1378
 QY 984 AATATCTATCTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1043
 DB 1379 GTCGACAACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1438
 QY 1044 AATTCTGTGCAAAATTTGAAGAAAACAAGCTCAGATTTATGATCGTGAATACTTCAAGAAG 1103
 DB 1439 GATTAAAAATCTCATTTGAGGACAGCAACAAGAAATATGAAAAGGAAAAGCTCAATGAGAG 1498
 QY 1104 TCTTGCNAACCTTTGTTGGTGGAGTAGCTGTATTCATGTTGGAGCTGCTACTGAACTGA 1163
 DB 1499 GATAGCAAAAGCTTTGCTGGTGGTGTGCTGCTCATTCAGTGGGAGCAACAACAGAACTGA 1558
 QY 1164 AATGAAGAGCAAGAGGATCGTGTAGAGATGCTCTAAATGCAACAAGAGCTGCGGTTGA 1223
 DB 1559 ACTAAAGAGAGAGAGTTAAGAGTTGAGGATGCCCTTAATGCAACCAAGGCTGCTGTTGA 1618
 QY 1224 AGAAGGTATTGTCCTGGTGGTACTGCTTTTGTCCGCTCCATTTAAAGTCTTGTGAAGA 1283
 DB 1619 GGAAGGTATTGTTGTTGGTGGAGGATGCACCCCTTTTGGCGCTTGCAGCTAAAGTCGATGC 1678
 QY 1284 TATTAA-----ACCTGCTGATGATGATGAACCTTGTGGAGCTTAATATCATCCGTCGTT 1337
 DB 1679 CATCAAGGATATACCTGGAGAACGATGAGCAGAAAGTTGGAGCTGAAATAGTAGAAGAGG 1738
 QY 1338 TCTTGAAGAGCTTTTACGTCAAAATGCTGCAAAATGCTGGCTATGAAAGTTCTATTGTTGT 1397
 DB 1739 CCTGAGCTACCCACTTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1798
 QY 1398 AGAAAAGTTTCGTG---AACCAAAAGATGTTTGGATTTTAAATGCTGCTCATCAGAGATA 1454
 DB 1799 TGAGAAGTCTCTTCTCAACGATTAATCAAGTATGTTTACAAATGCTCTACTTGACAGTA 1858
 QY 1455 TGAAGACCTTATTAAAGCTGTTGATTCCTTAAAGAGTTTACAGTATTGCAATTACA 1514
 DB 1859 CGAGGACTTGCATGGCTGCTGTTATCATTTGATCTCTACTTAAGTGGTGGAGTGTGTTTGA 1918
 QY 1515 AATGCAAGCTCAGTAGCCTCTTACTTCTTAACATACAGA 1553
 DB 1919 GCATGCCGCTGCGTGGCCCAAGACCTTCTCGACATCTGA 1957

AY106465 1376 bp mRNA linear HTC 16-OCT-2002

LOCUS
DEFINITION
Zea mays PC0128697 mRNA sequence.
ACCESSION
AY106465
VERSION
AY106465.1 GI:21209543
KEYWORDS
HTC.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1376)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1376)
Coe,E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES
source
1..1376
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="maizeDB:637486"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT
397 a 277 c 359 g 343 t

Query Match 23.9%; Score 394.4; DB 11; Length 1376;
Best Local Similarity 59.0%; Pred. No. 9.6e-82;
Matches 696; Conservative 0; Mismatches 481; Indels 3; Gaps 1;

448 ATTTCTGCAAACTCTGATACAACTAGTAAATATATCATAGCTGAAGCTATGCTAAAGTT 507
1 ATTTCTGCAAACTGAAAGAAATTTGAGATCTAATATCAAAAGCCATGAAAGTT 50
508 GGAAAGAGGTTGTTATCACAGTGGAGAGCTAAAGGCTTTGAAACTACATTAGATGTG 567
61 GGAAAGAGTGGAGTCAATTACTATTGTTGATGGCAAAACATTGGCAATGAGCTTGAAGCA 120
568 GTTGAGAGGATGAAGTTTACCGTGTACCTCTCCATATCTTTGTAATGAGAAAGAAATTAAGT 627
121 GTACAGGAATGAAGCTGTCAGAGGATACATATCTCTTACTTTGTGACTGATCAAAAG 180
628 AAAATGGTTTGTGAACCTTATATACCTTTGTAATGAGAAAGAAATTAAGT 687
181 ACTCAGAATGTGAGATGAGAACCTCTTATCTTATCCATGACAGAAATCTCAAC 240
688 ATGAAGACATGCTACCAATCTTAGAACAGGTTGCTTAAAGTAAACCGTCCACTCTTATT 747
241 ATGGAATCTCTCTCCAGCATTAGAAATTTCTATCAAGAAATGCGAAGCTCTTCTCAT 300
748 ATTGCTGAGAGCTGAGAGTGAACCTTGCACACCTTGATCAATAGCTCCGTGA 807
301 GTTCTGAGGATGTTGAAGGAGATGCTCTGTCAATGCTTGTACTGAAAGCATCGTGT 360

QY 808 GCACTCAAAGTTGTAGCCGCTAAAAGCTCCTGGTTTTTGGTGAACGCGTAAAGCTATGCTT 867
DB 361 GGACTCAAGGTATGTCTGTCAAGCTCTCTGATTTGGTGAATATAGAGGCAATCTA 420
QY 868 GAAGATATTGCTATCCTTACTCGAGGAGAGCAATATTTTGAAGATCGTGTATAAAGCTT 927
DB 421 GATGACATGGCTGTGATGACTCGAGGAGAGGTTATTAGCGAGGAACGCGTCTTGTATCT 480
QY 928 GAAATGTAAAGCTTGTCTTCTTTAGAACAGCTAAACGTGTAGTTATTGACAAAGAAAT 987
DB 481 GCGAAGTTCAATTAACAATGCTTGGCACTGTCTAAAGAGTTAACTGTATCCCTTGTATG 540
QY 988 ACTACTATCTGTTGATGCTCGAATAATCAGAAAGATATTAAAGCTCGAGTTAAACAAAT 1047
DB 541 ACTATCATCTAGATGGTGGAGTGACAAACAGCAGATAGATGAGGTTGCCACACAGCTT 600
QY 1048 CGTGCAAAATTTGAAGAAACAAGCTCAGATTATGATCGTGAATAAATTTCAAGAACCTCTT 1107
DB 601 AGAGAAATCAATTTGACACAAGCACTGCTGTATTTCGACAAAGAAAGGCCCAAGAGCTTTA 660
QY 1108 GCAAAATCTTGTGGTGGAGTAGCTGTATCCATGTTGGAGCTGCTACTGAAACTGAAATG 1167
DB 661 TCTAAGTTATCTGGAGGTTGCTGTCTACTCAAGATTGGTGGAGCTAGTGAAGCTGAAGTT 720
QY 1168 AAAGAGAGAGAGGATCGTGTAGAGATGCTCTAAATGCAACAGAGCTCGGTTGAAGAA 1227
DB 721 GGTGAGAAAGATAGAGGTGACAGATGCTTAATGCTGCAAGGGCTGCTGTGGAGGAG 780
QY 1228 GGTATGTCCCTGGTGGTACTGCTTTTGTCCGCTCCATTTAAAGTCCCTTGTATGATATT 1287
DB 781 GGCATTGTGCCAGGTGGTGGAGTTGCCCTTCTTTATGCCAACAGAGCTTGCACAGATT 840
QY 1288 ABACTGCTGATGATGAACTTCTGCACTTAATATCATCTCGTCTGTTCTTCTTGAAGAG 1347
DB 841 AGCAGCAAAACAGAGCAAGAAAGATTGGAGTTCAATTTATCAAGAAATTTCTTTGAAGCT 900
QY 1348 CTTTGTACGTTCAATTTGCTGCAATGCTGCTATGAAAGTTCTATTGTTGTAGAAAAGTT 1407
DB 901 CTTTGTATGATGATGCTGCAATGCTGCAATGCTGCAATGCTGCAATGCTGCAATGCTG 960
QY 1408 CGTGAACCAAAAGA---TGGTTTTGATTTAATGCTGATCAGGAGAGAAATATGAAGACCTT 1464
DB 961 ATTGAACAAGAGGACCTCAGTTTGGGCTATGATGACGCAAAAGGGAGTACGTAGACATG 1020
QY 1465 ATTAAGCTGCTGATGATGCTTAAAGAGTTACAGTATTGCTATGCAATGCAATGCAAGCA 1524
DB 1021 ATCAAGGCGGATCATCATGATCCCGTGAAGGTTGATCCGACCGCACTTCAAGACGCTGA 1080
QY 1525 TCAGTAGCTCTCTTACTTCTAACTACAGATGCTGCTATTGCTGAAAAACAGAACTTAA 1584
DB 1081 AGTGTTCCTGCTGATGCAACACAGGAGCGCTGTCTCCGAGCTCCCGCAACCAAG 1140
QY 1585 AAAGATATGCTTATGCTGGCGGTGTTGTTGGTGGTATGG 1624
DB 1141 TCCAGATAGCCAGCGGTATGCGGAGATGAGCGGATGG 1180

RESULT 7
LOCUS
AY108560
DEFINITION
Zea mays PC0140434 mRNA sequence.
ACCESSION
AY108560
VERSION
AY108560.1 GI:21211654
KEYWORDS
HTC.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2143)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 JOURNAL Overgo Probes
 REFERENCE 2 (bases 1 to 2143)
 AUTHORS Coe, E.H.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 COMMENT If you are interested in getting corresponding physical clones,
 these are publicly available from ZmDB and may be found by BLAST
 searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
 www.zmdb.iastate.edu.

FEATURES Location/Qualifiers
 source 1..2143
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="MaizeDB:638266"
 /clone_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 546 a 494 c 598 g 505 t
 ORIGIN

Query Match 23.9%; Score 394.4; DB 11; Length 2143;
 Best Local Similarity 54.5%; Pred. No. 1.2e-81;
 Matches 858; Conservative 0; Mismatches 706; Indels 10; Gaps 3;
 QY 8 CTAAGAAATCTTTTGTATTAAGCCGCTGAAACCTTTCAGAGGTGTAGATAAAC 67
 DB 275 CCAAGGAGATCGCTTTGACAGGGCTCCAGAGCGCCCTTCAGCGCGCGTCCAGAGC 334
 QY 68 TTGCAATGCTGTTAAAGTAACTTGGACCTTAAGCCGCTTAATGTCTGTTATTTGAAAGT 127
 DB 335 TCGCCGCGCGGTGGCGCTCACCTCGAGCCCAAGAGGAAGTGTCTTGGATGAG- 393
 QY 128 CTTTGTGTTCCCAAGTTATTACAAAGATGCTGTATCTGTGTGCAAAAGAAATTAAGCTTG 187
 DB 394 --TTTGAACCCCAAGTGTCTATGATGAGTTACCATTTGCCCGCGCTATTGAGCTTG 451
 QY 188 AAGATAAGTTTGAATPATGGCGCTCAAAATGTTTAAAGAGTAGCTCCCAAAACTAGCG 247
 DB 452 CTGATCCATGGAGATGCTGGTCTTCTTGTATCTGTGAAGTTGCTAGCAAGCGAATG 511
 QY 248 ATATTGCTGTGATGAAGTAC-AACAGCAAGTCTTGGCAAGCTATTATTCGGTAA 306
 DB 512 ACTCGGCTGTGTGTAAGCAACCAAACTGCTCTGTTCTCGCTAGGGAGATCATCAATTG 571
 QY 307 GGTGTAAACCTTCTAGCAGCTGTGTGTAATCTTATGCTTAAAGAGTAAAGTAAAGTAAAG 366
 DB 572 GGCATGTTGAGCAATCTTCAGGGCCAAATCCAGTGTCCGTGAAGGCAATGATAAG 631
 QY 367 GCTGTTGTTGCTGTACTAAAGAACTTAAGCGACATTAAGAGCTACTCTGTCACCAAAA 426
 DB 632 ACTGTTTCAAGAAATGGTGGAGGAATTTGAGAAGAAATCAAGCGCGTCAAGGGTAGTGGG 691
 QY 427 GAATAGCTCAAGTTGGAACCATTTCTGAAATCTTGATACAAATAGTAAATATATCAT 486
 DB 692 GATATTAAAGCTGTGTGTCATATCAGCTGGAATGATGAATTTGTTGGGACTATGATC 751
 QY 487 GCTGAAGCTATGCTAAAGTTGGAAGAGGTGTTATCACAGTTGAGGAGTAAAGGT 546
 DB 752 GCCAAGCTATTGACAGGTTGGCCCTGATGGTGTCTCTCAATTGAGTGTGTCATGCTG 811
 QY 547 CTTGAAACTACATTAGATGTGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606

AL559862 1201 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION AL559862 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CS0DG003YN06 5-PRIME, mRNA sequence.
ACCESSION
VERSION AL559862.2 GI:31283993
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12905763.
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1228.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DG003DG03QP1&cluster=1228.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DG003DG03QP1.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG003YN06"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 369 a 200 c 291 g 302 t 39 others
ORIGIN
Query Match 21.3%; Score 350.4; DB 9; Length 1201;
Best Local Similarity 57.4%; Pred. No. 2e-71;
Matches 619; Conservative 15; Mismatches 441; Indels 3; Gaps 1;
QY 218 TGGTTAAGAGTAGTCCCAAACTAGCGATATTGCTGCTGATGGAAGTACACACGAA 277
DB 23 TGTACAAAAGCAGCGTGTACCGGTCCCGAAATCCCGGATTCGGCATTACACTGCTA 82
QY 278 CAGTCCTTGACAGCTATTATCGTGAAGTGTAAAACTTTGACAGCTGGTGTATC 337
DB 83 CTGTACTGGCAGCTCTATACCAAGAGGCTTCGAGACATTTAGCAAGGTGCTAATC 142
QY 338 CTATGGCCATTAAACGTTGGCATAGATAAGCTGTGTTGCTGTTTACTTAAAGAACTAAGC 397
DB 143 CAGTGGAAATCAGGAGAGGTGTGATGTTAGCTGTGTTGCTGTTAATTCGTAACCTAAA 202
QY 398 ACATTACAGAGCTACTCGTGACCAAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAA 457
DB 203 AGCAGCTAAACCTGTGACCAACCCCTGAAGAAATTCACAGGTTGCTACGATTTCTGCAA 262
QY 458 ACTCTGATACAACTAGTAAATATCATGCTGAAGCTATGCTGTAAGTTGGAAGAGAG 517
DB 263 ACGAGACAAAGAAATTTGGCATATCACTCTGATGCAATGAAAGAAAGTTGGAAGAGG 322
QY 518 GTGTTATCACAGTTGAGAGCTTAAAGCTTTGAACTACATTAGATGTTGTTGAGGAA 577
DB 323 GTGTCAATCACAGTAAGAGTGGAAAAACACTGAATGATGAATTTAGAAATTTAGAGSCA 382

QY 578 TGAAGTTTGAACCGTGGCTACCTCTCCATACCTTTGTAACTAATCTCTGAGAAAATGGTTT 637
DB 383 TGAAGTTTGAACCGTGGCTACCTCTCCATACCTTTGTAACTAATCTCTGAGAAAATGGTTT 442
QY 638 GTGAACCTTGAACCGTGGCTACCTCTCCATACCTTTGTAACTAATCTCTGAGAAAATGGTTT 697
DB 443 GTGAACCTTGAACCGTGGCTACCTCTCCATACCTTTGTAACTAATCTCTGAGAAAATGGTTT 502
QY 698 TGCTACCAATCTTAGAACAAGTTGCTAAGTAGTAACCGTCCACTCTCTTATTTATTTGCTGAAG 757
DB 503 TTGTACCTGCTCTTGAATGCAATGCTCACCGTAAGCTTTGGTTCATAATCGTGAAG 562
QY 758 ACGTGAAGGTGAACGACCTTGAACGACCTTGTAGTCAATTAAGCTCCGTTGAGGACCTCCAG 817
DB 563 ATGTTGATGGAGAAGCTCTAAGTACACTCGTCTGAATAGGCTAAGAGTTGGTCTTCAGG 622
QY 818 TTGTAGCCGTAAAGCTCTGCTTTTGGTGAACCGCTAAAGCTATGCTTGAAGATATTG 877
DB 623 TTGTGGCAGTCAAGGCTCCAGGGTTTGGTGACAATAGAAAACACAGCTTAAAGATATG 682
QY 878 CTATCTTACTGGAGAGCAATATTTGAAGATCGTGTA---TAAAGCTTGAAGATG 934
DB 683 CTATCTTACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 742
QY 935 TAAGCTTGTCTTCTTGAAGACAGCTTAAAGCTGTAGTATTGACAAAGAAATACTACTA 994
DB 743 TTGAGCTCATGACTTAGGAAAAGTTGGAGAGTCAATGTCGACCAAGAGCGATGCGCATGC 802
QY 995 TCGTTGATGCTGCTGAAAATCAGAGATATTAAAGCTCGAGTTAAACAAATTCGTGCAC 1054
DB 803 TCTTAAAGGAAAAGGTGACAGGCTCAATTTGAAAACGATTTCAAGAAATCATTTGAGC 862
QY 1055 AATTCGAAGAAACAAGCTCAGATTATGATCGTGTAAGAAACCTTCAAGAACGCTTTGCAAAAC 1114
DB 863 AGTTAGATGTCACAACTAGTGATATGAAAAGGAAAACCTGAAATGAACGCTTTGCAAAAC 922
QY 1115 TTGTTGG 1174
DB 923 TTTGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 982
QY 1175 AGAGATGCTGTAGAGATGCTCTAAATGCAACAAGAGCTGCGGTTGAAAGAGGATTTG 1234
DB 983 ARAAGACAGATTACAGATGCCCTTATGCTACAGAGCTGCTGCTGAGGAGGCTTTG 1042
QY 1235 TCCTG 1292
DB 1043 TTTTGGGGGGGGTGGGGGGTGGGGGGTGGGGGGTGGGGGGTGGGGGGTGGGGGGTGGGGGGTGGGG 1100

RESULT 9
BM799922
LOCUS
DEFINITION BM799922 1090 bp mRNA linear EST 05-MAR-2002
AGENCOURT_6417164 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5492101
5', mRNA sequence.
ACCESSION
VERSION BM799922.1 GI:19116745
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

http://image.llnl.gov
Plate: LLAM12113 row: k column: 14
High quality sequence stop: 743.
Location/Qualifiers
1..1090
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5492101"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: PCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
FEATURES
source

```

BASE COUNT	332 a	209 c	274 g	275 t	ORIGIN
Query Match	19.6%	Score 323.4	DB 12	Length 1090	
Best Local Similarity	59.4%	Pred. No. 4.3e-65			
Matches 549	Conservative 0	Mismatches 376	Indels 0	Gaps 0	
QY	6	TTCTAAAGAAATCC	TTTTTGATGCTAAAGCCCGTGA	AAAACTTTCACGAGGTGTAGATAA	65
Db	121	TGCCAAAGATGTAA	AAATTTGGTGAGATGCCCGAGCCTT	AATGCTTCAAGGTGTAGACCT	180
QY	66	ACTTGGCAATGCTCT	TAAAGTAACACATTGGACCTTAAAGCGCGTA	ATCTCGTATTTCGAAA	125
Db	181	TTTAGCCGATGCTGT	GCGGTACAATTGGGGCCAAAGGGAAGAACAGT	GAATTTATGACGA	240
QY	126	GTCCTTTTGGTTC	CCCCAGTTATTACAAAGATGGTGTATCT	TTCGAAAAGAAATTTGAAC	185
Db	241	GAGTTGGGGAAGT	CCCCAAGTAACAAAGATGGTGTGACT	TTGTTCGAAAGTCAATTTGACT	300
QY	186	TGAAGATAAGTTT	CGAAATATGGCGCTCAAA	TGGTTTAAAGAGTACTCCCAAACTAG	245
Db	301	AAAAGATAAAT	ACNAGAACATTTGGAGCTAACT	TTGTTTCAAGATGTTGCCAATTAACACAA	360
QY	246	CGATATTGCTGGT	GATGGAACTACAAACAGCAACAGTCC	TTGCACAAAGCTATTTATCGTGA	305
Db	361	TGAAGAGCTGGG	ATGSCACTACCTGCTACTGTACTGTG	CGACGCTCTATAGCCAAAGGA	420
QY	306	AGGTGTAAACTT	GTGTAGACGCTGGTGTATTCCTATGCGCAT	TAAACGTGGCATAGATAA	365
Db	421	AGGCTTCGAGA	AGATTTAGCAAGGTGCTAATCCAGTGGAA	ATCAGGAGAGGTGTGATGTT	480
QY	366	AGCTGTTGTGTCT	TTACTAAAGAACTTAAGCGCATTTACAAAG	CGCTACTCGTGACCAAAA	425
Db	481	AGCTGTGTAGCT	GTAAATGCTGNACTTAAACAGCAGTCT	TAAACCTGTGACCCCTCTGA	540
QY	426	AGAAATAGCTCA	AGTTGGAAACCAATTCGCGAACTCT	GATACACAATAGGTAAATCAT	485
Db	541	AGAAATGCA	AGGTTGCTACGATTTCTGCAACCGGAGACA	AAAAATTTGCGCAATATCAT	600
QY	486	AGCTGAAGCTAT	CGCTAAAGTTGGAAAGGAGGTGTTAT	CACAGTTGAGGAAGCTAAAGG	545
Db	601	CTCTGATGCA	ATGAAAAGTTGGAAGAAGGTTGTCAT	CACAGTAAGGATGGAANAAC	660
QY	546	TCTTGAACAT	TACATTAGATGTGGTTGAAGGAATGAAG	TTTTGACCGTGGCTACCTCTCTCC	605
Db	661	ACTGAATGAT	GCAATTAGAAATTTATTGAAGGCAT	GAACTTTGATCGAGGCTATATTTCTCC	720
QY	606	ATACCTTTGTAA	CTAATCCTGAGAAAATGTTTTGTGAAC	TTTGATTAACCCCTTATATCCTTTG	665
Db	721	ATACCTTTAT	TAAATACATCAAAAGGTACAAATGTGA	ATTCGAGATGCCCTATGTTCTGTT	780
QY	666	TAATGAGAAAAA	AGATTACTAGCATGAAGACATGCTAC	CAATCTTAGACAAGTTTGCTAA	725
Db	781	GAGTGAAAAG	AAAATTTCTAGTATCCAGTCCATTG	TACCTGCTCTTGAAATTTGCCCATGC	840
QY	726	AGTAAACCGT	CCACTCCCTTATTTATGCTGAAGACGT	TAGAAGGTGAAGCACTTGCAACACT	785

Db	841	TCACCGTAAAGCCTTTGGTCAATAATCGCTGAAGANGTTGTATGGAGAAAGCTCTAAGTACACT	900
Qy	786	TGTAGTCAATAAGCTCCGTGGAGCACTCCCAAGTTGTAGCCGTAAAAGCTCCTCGTTTGG	845
Db	901	CGTCTTGAATAGGGCTAAAGGTGGTCTCCAGGGTGTGGCACTCAAGGCTCAGGTTTTGG	960
Qy	846	TGAACGCCGTAAAGCTATGCTTGAAGATATGCTATCTCTTCTTCTGGAGGAGAAACAATATT	905
Db	961	TGACATAGAGAAACCCAGCCTAAAGATATGGCTATTGCTACTGGTGGGCGCAAGGTTTT	1020
Qy	906	TGAAGATCGTGGTATAAAGCTTGAA	930
Db	1021	GGAAAAAAGGGATTGACCCTTGAA	1045
RESULT 10			
LOCUS	BH770613/c		
DEFINITION	LMWGag372 MG1363 Random sequence Tag Library <i>Lactococcus lactis</i> subsp. <i>cremoris</i> genomic, genomic survey sequence.	1327 bp DNA linear	GSS 01-MAY-2002
ACCESSION	BH770613		
VERSION	BH770613.1	GI:20373570	
KEYWORDS	GSS.		
SOURCE	<i>Lactococcus lactis</i> subsp. <i>cremoris</i>		
ORGANISM	<i>Lactococcus lactis</i> subsp. <i>cremoris</i>		
REFERENCE	1 (bases 1 to 1327)		
AUTHORS	Bolotin,A., Ehrlich,S.D. and Sorokin,A.		
TITLE	Studies of genomes of dairy bacteria <i>Lactococcus lactis</i> Sgl. Aliments, (2002) In press		
JOURNAL	Contact: Sorokin A		
COMMENT	Genetique Microbienne INRA CPI INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France Tel: 33 1 34 65 25 16 Fax: 33 1 34 65 25 21 Email: sorokine@jouy.inra.fr best homologue in strain IL1403 is groEL (93%) Class: shotgun High quality sequence start: 30 High quality sequence stop: 1299.		
FEATURES	source		
	1..1327		
	/organism="Lactococcus lactis subsp. cremoris"		
	/mol_type="genomic DNA"		
	/strain="MG1363"		
	/db_xref="taxon:1359"		
	/clone_lib="MG1363 Random Sequence Tag Library"		
	/notes="Vector: pSGM2; Site 1: SmaI; Library of chromosomal fragments of <i>L.lactis</i> strain MG1363 was prepared by partial AluI digestion or by sonication."		
BASE COUNT	392 a 292 c 239 g 403 t		
ORIGIN			
	Query Match	19.5%;	Score 321; DB 28; Length 1327;
	Best Local Similarity	65.7%;	Pred. No. 1..7e-64;
	Matches	498; Conservative	0; Mismatches 256; Indels 4; Gaps 2;
Qy	823	GC CGTAAAGCTCTGTTTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATGCTATC	882
Db	1321	GCTGTAAAGACACGGGATTGGTG-CCGCCGTAAAGCTCAATTGGAAGATTAGCTATC	1263
Qy	883	CTTACTGGAGAGAACCAATATTGAGATCGTGTATAAGCTTCAAAATGTAAGCTTG	942
Db	1262	TTGACAGTGGTACAGTAATACCTGTTGATTAAGACCAATGCTACACTA	1203
Qy	943	TCTTCTTTAGAACACAGCTAAACCGCTGTAGTTATTGACAAAGAAATACTACTATCGTTGAT	1002
Db	1202	GAGCTTTAGCAACAGCTGCTAAAGCATCTGTTGATTAAGACCAATGCTTGAA	1143
Qy	1003	GGTGCTGGAAAAATCAGAAGATATAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAA	1062

QY 1386 TTCTATTGTTGTAGAAAAGTTTCGTGAACCAAAAAGATGG-TTTTGGATTAAATGCTGCAT 1444
 Db |||||
 QY 846 TGAAGTGTGTTGTCAGAGAGATCAAGGCAGACGATTTGGAGGTGGGCTACATCAATGA 905
 Db |||||
 QY 1445 CAGAGAAATATGAAGACTTATTAAAGCTGGTGTCAATGTATGATCTTAAAGAAAGTTACACGTA 1504
 Db |||||
 QY 906 CTGACAAAGTACGAGAACTGATGAGGCTGGTGTGATCGACCTTGCACAAAGTGACCAAGT 965
 Db |||||
 QY 1505 TTGCAATTAACAAATGACGATCACTAGCTCTTACTTCTAACTACAGATGCGCTATTG 1564
 Db |||||
 QY 966 GTGGCTCCAGATGCGGCTCTAGTAGCTGAATGGTCTTAACACGAGGCCATTTGTTG 1025
 Db |||||
 QY 1565 CTGAAAACACAGAACCTTAAAGAAATATGCTATGCC 1601
 Db |||||
 QY 1026 TTGAGAAGCCAAAGCCGAGCTCAAGTGGCAGAGCC 1062
 Db |||||

RESULT 12
 AL515262
 LOCUS
 DEFINITION AL515262 1172 bp mRNA linear EST 08-MAY-2003
 CL0B016ZA04 5-PRIME, mRNA sequence.
 ACCESSION AL515262
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1172)
 Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished
 On Feb 13, 2001 this sequence version replaced gi:12778755.
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 1228.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CL0B016ZA04RPI&cluster=1228.r. Contact :
 Feng Liang Email: fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ invitrogen Corporation 1600
 Paraday Avenue Genoscope sequence ID : CL0B016ZA04RPI.

FEATURES
 source
 1. .1172
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CL0B016ZA04"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched
 double-strand cDNA was digested with NotI and cloned into
 the NotI and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."
 BASE COUNT 352 a 216 c 293 g 294 t 17 others
 ORIGIN
 Query March 19.4%; Score 318.8; DB 9; Length 1172;
 Best Local Similarity 59.9%; Pred. No. 5.4e-64;
 Matches 543; Conservative 4; Mismatches 357; Indels 2; Gaps 1;
 QY 6 TTCTAAAGAAATCCTTTTTCATGCTAAAGCCCGTGAAGAACTTTTCAGAGGTGTAGATAA 65
 Db |||||
 QY 164 TGCAGAGATGTAATTTGGTCAGATGCCGAGCCCTTAATGCTTCAAGGTGTAGACCT 223
 Db |||||
 QY 66 ACTTGCAATGCTGTTTAAAGTAACACTTGGACCTAAAGCCGTAATGTGCTGTTATTGAAA 125
 Db |||||
 QY 224 TTTAGCGATGCTGTGGCCGTTTCAATGCGGCCCAAGGAAGACAGTGATTATTGACGA 283
 Db |||||

QY 126 GTCTTTTGGTTCCTCCAGTTATTACAAAGATGGTGTATCTGTTGCAAAAGAAATTTGAAT 185
 Db |||||
 QY 284 GAGTTGGGAAGTCCCAAGTAAACAAAGATGGTGTGACTKTTGCAAGTCAATTTGACTT 343
 Db |||||
 QY 186 TGAAGATAAGTTTGAATAATGGGCGCTCAATGGTTAAAGAACTAGTCTCCCAAACTAG 245
 Db |||||
 QY 344 AAAAGATAAATACAAAAACATTGGAGCTAAACTTTGTTCAAGATGTTGCCAATAACAAA 403
 Db |||||
 QY 246 CGATATTGCTGGTGAAGAACTACAAAGCAACAGTCTCTTGCACAAGCTATTATTTCGTGA 305
 Db |||||
 QY 404 TGAAGAAAGCTGGGAGTGGCACTACCACTGCTACTGTATGTCGACGCTCTATAGCAAGA 463
 Db |||||
 QY 306 AGGTGTAAACTTGTAGCAGTGTCTTAATCTTATGCCATTAACAACTGTCGATAGATAA 365
 Db |||||
 QY 464 AGCTTCGAGAAGATTAGCAAGGTGCTAATCCAGTGAATCAGGAGAGGTGTGATGTT 523
 Db |||||
 QY 366 AGCTGTTGTTGCTGTTACTAAAGAACTAAGCACTTACAAAGCTTACTCGTGACCAAAA 425
 Db |||||
 QY 524 AGCTGTTGATGCTGTTATTGCTGAACCTTAAAGCAGCTTTAAACCTGTACCACCCCTGA 583
 Db |||||
 QY 426 AGAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGTATACAACTAGTAAATATCAT 485
 Db |||||
 QY 584 AGAATATGCAACAGTTGCTACGATTTCTGCAAAACGAGACAAAGAAATTTGCAATATCAT 643
 Db |||||
 QY 486 AGCTGAGCTATGGTAAAGTTTGGAAAGAGGTGTTATCACAGTTGAGGAAGTAAAGG 545
 Db |||||
 QY 644 CTCGTGATGCAATGAAAAAGTTTGGAAAGAGGTGTCATCACAGTTAAGGATGGAAAAAC 703
 Db |||||
 QY 546 TCTTTGAAACTACATTTAGATGTTGAAAGAAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
 Db |||||
 QY 704 ACTGAATGATCAATTAGAAATTTATGAGGCATGAAGTTTGATCGAGGCTATTCTCTCC 763
 Db |||||
 QY 606 ATACTTTGTAACCTCTCGAGAAATGTTTGTGAACCTTGTAACCTTATATFCCTTTG 665
 Db |||||
 QY 764 ATACTTTATTATATACATCAAAAGCTCAGAAATGTGAATCCAGGATGCTATGTTCTGTT 823
 Db |||||
 QY 666 TAATGAGAAAAAGTTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTTGCTAA 725
 Db |||||
 QY 824 GAGTGAAGAAAAATTTCTAGTATCCAGTCCATTTGACCTCTTGAGGKTGCCAATGC 883
 Db |||||
 QY 726 AGTAAACCGTCCACTCTCTATTATGCTGAAGACGTAGAGGTGAAGCACTTGCACACT 785
 Db |||||
 QY 884 TCACCGTAAGCCCTTTGGTCATATCGCTGAAGATGTTGATGGAGAGCTCTAAGTACACT 943
 Db |||||
 QY 786 TGTAGTCAATAAGCTCCGTTGAGCACTTCCAAAGTTGAGCCGTAAAGCTCCTCGTTTTGG 845
 Db |||||
 QY 944 CGTCTTGAATAGGTAAAGGTTGTTCTTCAAGTTTGGCAGTGAAGCTCCAGGGTTTG 1003
 Db |||||
 QY 846 TGAACCGCTTAAAGCTATGTTGAAGATATTGCTTACTTCTTACTGGAGGAGCAATATT 905
 Db |||||
 QY 1004 TGAM--ATGAAAGAACARCTTAAGATATGGCTATTGCTACTTGGTGGTGCAGTTTGG 1061
 Db |||||
 QY 906 TGAAGA 911
 Db |||||
 QY 1062 RVAAGA 1067
 Db |||||

RESULT 13
 AL518632
 LOCUS
 DEFINITION AL518632 1201 bp mRNA linear EST 12-MAY-2003
 CS00A0091115 5-PRIME, mRNA sequence.
 ACCESSION AL518632
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1201)
 Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 TITLE

JOURNAL
COMMENT

Unpublished
On Feb 13, 2001 this sequence version replaced gi:12782125.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1228.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODA009AE08QPl&cluster=1228.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODA009AE08QPl.

FEATURES

source

Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODA009Y115"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pcMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pcMVSPORT 6 vector.
Library was not normalized."

BASE COUNT 364 a 222 c 301 g 295 t 19 others
ORIGIN

Query Match 19.1%; Score 314.8; DB 9; Length 1201;
Best Local Similarity 58.4%; Pred. No. 4.7e-63;
Matches 541; Conservative 4; Mismatches 381; Indels 0; Gaps 0;

QY 6 TTCTAAGAGAAATCCCTTTTGTGATGCTTAAAGCCGCTGAAAGAACTTTCAAGAGGTGTAGATAA 65
DB 189 TGCCAAAGATGTAAATTTGGTGCAGATGCCCCGAGCCTTAATGCTTCAAGGTGTAGACCT 248
QY 56 ACTTCAATGCTGTAAAGTAACACTGGACCTAAGCCGCTAATGCTGTTATTGMAAA 125
DB 249 TTATGCCGATGCTGTGGCGGTTCACAAATGGGGCCAAAGGGAACAGTATTATTGAGCA 308
QY 126 GTCTTTTGGTCCCGAGTATTACAAAAGATGGTGTATCTGTTGCAAAAAGAAATTAAGACT 185
DB 309 GAGTTGGGAGTCCCAAGTAACAAAGATGGTGTGACTGTTCGCAAGTCAATTGACTT 368
QY 186 TGAAGATAGTTGAAAATATCGGCGCTCAATATGTTAAAGAGTAGCTCCCAAAACTAG 245
DB 369 AAAAGATAAATACAAAACCAATGGAGCTAAACTTTGTTCAAGATGTTGCCAATAACACAAA 428
QY 246 CGATATTGCTGTGATGGAACCTACACAGCAACAGTCCTTGCACAGCTATTATTATCGTGA 305
DB 429 TGAAGAGCTGGGATGGGACCTACCACCTGCTACTGTGACCTGGCAGCGCTTATAGCCAGGA 488
QY 306 AGGTGTAAAACTTTGACGAGCTGGTGTATATCTATGGCCATTAAACGTTGGCATAGATAA 365
DB 489 AGGCTTCGAGAGATAGCAAGGTGCTAATCCAGTGGAAATCAGGAGAGGTGTGATGT 548
QY 366 AGCTGTGTTGCTGTACTTAAAGAACTAAGCGACATTAACAAGCCTACTCGTACCAAAA 425
DB 549 AGCTGTGATGCTGTAATTTGCTGAACCTTAAAGAGCAGTCTAAACCTGTGACCCACCCCTGA 608
QY 426 AGAATAGCTCAAGTTGGAAACCATTTCTCAAACTCTGATACAACTAAGTAAATATCAT 485
DB 609 AGAAATTGCACAGGTGTCTACGATTTCTGCAACCGGAGCAAGAAATTTGGCAATATCAT 668
QY 486 AGCTGAAGCTATGCTTAAAGTTGGAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGG 545
DB 669 CTCTGATGCAATGAAAAGAGTTGGAAGAAAGGCTGTCATCAGATTAAGAGGATGGAAGAAC 728
QY 546 TCTTGAACATCAATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACTCTCTCC 605
DB 729 ACTGAATGATGAATTAGAAATTAATTGAAGGCATSCCGTTTGTATCGAGGCTATATTTCTCC 788

QY 606 ATACTTTGTAACCTAATCTCTGAGAAAATGGTTTGTGAACCTTGATAACCTTTATATCTTTG 665
DB 789 ATACTTTTAAATATACATCAAAAGGTCAGAAATGGAATTCAGGATGCTATGTTCTGTT 848
QY 666 TAATGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAAGTTGCTAA 725
DB 849 GAGTGAAAAGAAAATTTCTAGTATCCAGTCCCATTTGACCTGCTCTTGAATTTGCCAATGC 908
QY 726 AGTAACCGCTCCACTCCCTTATTATGCTGAACACCTAGAGCTGAAGCACTTGCAACACT 785
DB 909 TCACCGTAAAGCTTTGGTCATAATCGCTGAAGATTTGATGGAAGCTCTAAGTACACT 968
QY 786 TGTAGTCAATAAGCTCCCGTGGAGCACTCCAAAGTTGTAGCCGTAAAGAGCTCTCTGTTTGG 845
DB 969 CGTCTTGAATAGGCTAAAGGTTGGTCTTCAGGTTGTGGCAGTCAAGGCTCCAGGGGTTG 1028
QY 846 TGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCTCTTACTGAGGAGAAAGCAATATT 905
DB 1029 GTGACATAGAYAGAMCAGCTTAAAGATATGCTATTGCTACTGCTGCTGCTGCTGTTGGA 1088
QY 906 TCAAGATCGTGTATATAAAGCTTGAAA 931
DB 1089 GAGAGGATGACCTGAATCTKAGA 1114

RESULT 14

EX439563
LOCUS
DEFINITION
BX439563 Homo sapiens PLACENTA Homo sapiens cDNA clone CSODE012YCL1
5-PRIME, mRNA sequence.
ACCESSION
BX439563
VERSION
BX439563.1 GI:30773785
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1228.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODE012AB06QPl&cluster=1228.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODE012AB06QPl.

FEATURES

source

Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODE012YCL1"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pcMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pcMVSPORT 6 vector.
Library was not normalized."

BASE COUNT 348 a 227 c 290 g 46 others
ORIGIN

Query Match 19.0%; Score 312.6; DB 13; Length 1201;
Best Local Similarity 58.0%; Pred. No. 1.6e-62;
Matches 513; Conservative 18; Mismatches 353; Indels 0; Gaps 0;

Qy	1002	TGGTGTGGAAATCAGAGATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATGA	1061
Db	557	TGANGCAGCCTCAAGGATGAGATACANACTAGGTGGCACAACCTAAAAAAGAGCTAGC	616
Qy	1062	AGAAACAGCTCAGATTATGATCGTGAATAAACTTCAAGAACGTCTTGCAAAACTTGTGG	1121
Db	617	TCAGACAGATTCTGTCTATGATTCAGAAAAATTCAGAAAGGATTGCCAACTATCTSG	676
Qy	1122	TGGAGTAGCTGTATCCATGTTGGAGCTGCTACTGAACTGAATGAAGAGAGAAGGA	1181
Db	677	TGGTGTTCACATCAATTAAGGTGGGGCTGCAACAGAGACTGAACCTGAGGATCGTAAGCT	736
Qy	1182	TCGTGTAGAAGATGCTCTAAATGCAACAGAGCTCGGTGGAAGAGGTATTGTCCCTSG	1241
Db	737	ACGATTCGAAGATGCAAGATGCTACATTTGCTGCCATAGNAGAGGTATTGTGCTGG	796
Qy	1242	TGGTGGTACTGCTTTTGTCC	1261
Db	797	TGGTGTGCTGCTTAGTTC	816

Search completed: January 29, 2004, 15:25:33
Job time : 2474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 14:44:13 ; Search time 475 Seconds

(without alignments)
12637.020 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 1647

Sequence: 1 atggcttcaagaatactct.....gtatgacggtatgtactag 1647

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	739.4	44.9	1635	13 US-10-228-167A-1	Sequence 1, Appli
2	725.8	44.1	1830121	15 US-10-329-960-1	Sequence 1, Appli
3	718.2	43.6	1626	12 US-10-369-493-33982	Sequence 33982, A
4	710	43.1	640681	10 US-09-790-988-1	Sequence 1, Appli
5	701.4	42.6	1635	12 US-10-369-493-40819	Sequence 40819, A
6	698.4	42.4	3625	10 US-09-070-927A-42	Sequence 42, Appl
7	690.2	41.9	1230025	12 US-10-289-762-1	Sequence 1, Appli
8	689.8	41.9	1926	15 US-10-267-311-50	Sequence 50, Appl
9	689.6	41.9	1635	9 US-09-841-132-380	Sequence 380, App
10	684.4	41.6	2155	10 US-09-960-428-13	Sequence 13, Appl
11	684.4	41.6	3840	12 US-10-157-317-26	Sequence 26, Appl
12	684.4	41.6	3840	12 US-10-157-339-26	Sequence 26, Appl
13	684.4	41.6	3840	13 US-10-157-305A-26	Sequence 26, Appl
14	684.4	41.6	3840	13 US-10-157-391-26	Sequence 26, Appl
15	684.4	41.6	3840	13 US-10-157-096-26	Sequence 26, Appl

16	684.4	41.6	3840	13 US-10-157-215A-26	Sequence 26, Appl
17	684.4	41.6	3840	13 US-10-157-302-26	Sequence 26, Appl
18	684.4	41.6	3840	13 US-10-154-951B-26	Sequence 26, Appl
19	684.4	41.6	3840	13 US-10-157-299-26	Sequence 26, Appl
20	684.4	41.6	3840	13 US-10-156-831-26	Sequence 26, Appl
21	684.4	41.6	3840	13 US-10-156-902-26	Sequence 26, Appl
22	684.4	41.6	3840	13 US-10-157-147-26	Sequence 26, Appl
23	684.4	41.6	3840	13 US-10-157-166-26	Sequence 26, Appl
24	684.4	41.6	3840	13 US-10-157-318-26	Sequence 26, Appl
25	684.4	41.6	3840	13 US-10-156-811-26	Sequence 26, Appl
26	684.4	41.6	3840	13 US-10-157-320A-26	Sequence 26, Appl
27	684.4	41.6	3840	13 US-10-157-418A-26	Sequence 26, Appl
28	683.8	41.5	1635	12 US-10-369-493-46781	Sequence 46781, A
29	681	41.3	1641	12 US-10-369-493-33701	Sequence 33701, A
30	680.8	41.3	1620	12 US-10-369-493-23868	Sequence 23868, A
31	643.8	39.1	1644	12 US-10-369-493-41211	Sequence 41211, A
32	639.2	38.8	1724	13 US-09-882-227-417	Sequence 417, App
33	637.4	38.7	1623	12 US-10-369-493-32974	Sequence 32974, A
34	629.2	38.2	1736	13 US-09-841-260-19	Sequence 19, Appl
35	629.2	38.2	1736	14 US-10-007-693-19	Sequence 19, Appl
36	623.2	37.8	1638	12 US-10-369-493-23854	Sequence 23854, A
37	607.6	36.9	1632	12 US-10-369-493-35447	Sequence 35447, A
38	604.4	36.7	1629	12 US-10-369-493-38952	Sequence 38952, A
39	587.4	35.7	1653	12 US-10-369-493-43871	Sequence 43871, A
40	585.2	35.5	1647	12 US-10-369-493-34549	Sequence 34549, A
41	563.4	34.2	1653	12 US-10-369-493-42681	Sequence 42681, A
42	560	34.0	1632	12 US-10-369-493-39155	Sequence 39155, A
43	560	34.0	1632	12 US-10-369-493-39904	Sequence 39904, A
44	553.6	33.6	1704	13 US-10-032-585-6410	Sequence 6410, Ap
45	551.8	33.5	1620	12 US-10-369-493-39525	Sequence 39525, A

ALIGNMENTS

RESULT 1

US-10-228-167A-1

; Sequence 1, Application US/10228167A

; Publication No. US20030147909A1

; GENERAL INFORMATION:

; APPLICANT: Hernan Marshall Gonzalez, Sergio

; TITLE OF INVENTION: HIGHLY IMMUNOGENIC PROTEIN AGAINST THE INTRACELLULAR PATHOGEN AGER

; TITLE OF INVENTION: PISCIRICKETTSIA SALMONIS, WHICH AFFECTS SALMON CULTURE, AMINO ACI

; TITLE OF INVENTION: ACID SEQUENCES OF SAID PROTEIN AND ITS APPLICATION IN THE DEVELOI

; TITLE OF INVENTION: METHODS FOR THE PREVENTION AND DIAGNOSIS OF DISEASES CAUSED BY SI

; FILE REFERENCE: 076502-9004

; CURRENT APPLICATION NUMBER: US/10/228,167A

; CURRENT FILING DATE: 2002-08-26

; PRIOR APPLICATION NUMBER: Chilean Patent Application No. US20030147909A1 2086-2001

; PRIOR FILING DATE: 2001-08-27

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 1

; LENGTH: 1635

; TYPE: DNA

; ORGANISM: Piscirickettsia salmonis

US-10-228-167A-1

Query Match 44.9%; Score 739.4; DB 13; Length 1635;
Best Local Similarity 66.2%; Pred. No. 2.7e-141;
Matches 1086; Conservative 0; Mismatches 546; Indels 9; Gaps 1;

Qy	1	ATGCGTCTTAAGAAATCCCTTTTGATGCTAAAGCCGCTGAAACCTTTCACGAGGTGTA	60
Db	1	ATGTCACGAAAAGAGTGGCGCTTCGGTACCGTTCCCGTCAAAAATGTTGACGCGGTGT	60
Qy	61	GATAAACTTGCAAATGCTGTTTAAAGTAACTTCGACCTAAGGCCGCTTAATGTCGTTATT	120
Db	61	AACCTACTTGCAAATGAGTGAAGTCACTCTAGTCCACGCGACGCAAGCTTATTITA	120
Qy	121	GAAAAGCTTTTGGTCCCGAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT	180
Db	121	GAAAAGTCATTTGGTGGCCCCCAACCATCACTAAAGATGGTGTATCTGTTGCCAAAGAAATC	180

181 GAACCTTGAAGATAGTTTCAAAATATGGCGCTCAAAATGGTTAAAGAGTAGCTCCAAA 240
181 GAGCTTAGCGATAGTTTCAAAACATAGGCGGACAAATGGTCAAGAGTGGCATCTAAA 240
241 ACTAGCGATTTAGTGGTGGTGAAGTACAAACAGCAACAGTCCCTTGCACAAGCTATTTAT 300
241 TCAAAATGATGATGACGGTGAACGATACGACGCGGACAGTATTAGCACAAGCAATTTT 300
301 CGTGAAGGTGAAGCTTGTAGCAGCTGGTGGTAAATCTTATGCTTATGCTTAAACAGTGCATA 360
301 CAAGAAGGCGTGAAGTCTGTGTCGCGCATGAACCCAAATGACCTTAAACGCGGCATC 360
361 GATAAAGCTGTTGTTGCTGTTACTTAAAGAACTAAGCGACATTAACAAAGCCTACTCGTAC 420
361 GATAAAGCACTATCGTGCAGTTGCTGCTTAAAGACTTATCTACCGGTCACAGAC 420
421 CAAAAGAAATAGCTCAAGTTGGAACATTTCTGCAAACTCTGATACAACAATAGTAAAT 480
421 AACAAAGCCATTGCTCAAGTCGCTACAAATTCAGCAAACTCTGATGAAGAAATTTGCTCT 480
481 ATCATAGCTGAAGCTATGCTTAAAGTTGAAAGGAAATGAAGTTTTCACCGTGGCTACCTC 540
481 ATCATGCTTAAAGCGATGAAAGAAAGTACCTACCGACGCGCTAATCACTGTTGAAGAGGC 540
541 AAAGGCTTTGAACTACATTAGATGTGTTGAAGGAATGAAGTTTTCACCGTGGCTACCTC 600
541 TCCAGCTTGAAGCAAGATTAAGTGTGTTGAAGGATGAAGTTTTCACCGTGGCTACCTC 600
601 TCTCCATCTTTGCTCAACAAACAGAGAAATGATCGCTGAAATCGAAAGCCCATTTATC 660
661 CTTTGTAAAGAAAGATTTACTAGCATGAAGACATGCTACCATCTTAGAACAAGTT 720
661 TTACTGCTGCAAGAAATTTCTAACATTCGCAATTTACTACCCCATTTAGAAATCAGTT 720
721 GCTAAAGTAAACCGTCCACTCTCTTATTGTTGGAAGCGTAGAAGTGAAGCACTTGCA 780
721 GCTAAATCAGCAAGCACTTATTCATCATCGTGAAGATTTGAAGTGAAGCTTGGCA 780
781 ACATGTTAGTCAATAGCTCCGTGAGCACTCCAAAGTTGTAGCGGTAAAGCTCTGTGT 840
781 ACATGCTGCTGTTAAATCAATTCGCGGTATTGTTAAAGTGTGCGCAGTAAAGCACTGSC 840
841 TTTGTTGAACGCGTAAAGCTATGCTGAAGATTTGCTTATGCTTACTGAGGAGCA 900
841 TTTGTTGATCGTGAAGCGATGCTTGAAGATTTGCTTATGCTTACTGAGGAGCA 900
901 ATATTGAAGATCGTGGTATAAAGCTTGAAGATTTGCTTCTTTTAGGAACAGCT 960
901 ATCTCTGAAGAGTTGGCTAGACCTTGAAGAGCACTCTTGAGCACTTAGTACAGCA 960
961 AAACGTGTATTGATCAAGAAATATCTATCTGTTGATGGTGGTCTGAAATCAGAA 1020
961 AAACGCATCGTCTCACTAAACACAATAACAACGTTTATTGATGTTGCGGTGAACAAAT 1020
1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATAT 1080
1021 GCGATCGAGCTCGGTTACTCAAAATCCGTGCACAAATTTGAAGAAACATCTCTGACTAC 1080
1081 GATCTGAAAAACTTCAAGAAAGCTCTTGAAGAACTTGTGTTGGAGTAGCTGTATCCAT 1140
1081 GACCCGGAAGAACTGCAAGAGGCTGCTGCTAGCTATCTGGTGGTGGTCTGCTCATTA 1140
1141 GTTGAAGCTGCTACTGAAACTGAAATGAAGAGAGAGGATCGGTGAGAAAGCTCTTA 1200
1141 GTTGGCGCAGCACTGAAATCAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
1201 AATGCAACAGAGCTGCGGTGCAAGAGGATTTGCTCCCTGTTGTTGTTGTTGTTGTTG 1260
1201 CATGCAACAGCGCGCTGCTGAGAGAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260

1261 CGTCTCAATTAAGCTCTTGATGATATTAACCTGCTGATGATGAACCTTGCTGACTT 1320
1261 CGTGAATCGCTGCACTTAAAGCTCTTGACTTCGAAATGATGAACCAAGCTGCT 1320
1321 AATATCATCGCTGCTTCTTGAAGAGCCTTTACGTTCAAAATGCTGCAAAATGCTGCTAT 1380
1321 AACATCTTGTGCTGCTATGAGCGCACCATTAACGTTCAAAATGCTGGAACGCGAGTGC 1380
1381 GAAGGTTCTATTGTTGTAGAAAAGTTCTGTAACCAAAAGATGTTTGGATTTATGCT 1440
1381 GAAGCGGCTGTAATCTTGTATAAAATGTCAACGCTGAAGGTAACCTTTGTTATAATGA 1440
1441 GCATCAGAGAAATATGAAGACCTTATTAAGCTGCTGCTCAATGATCCTAAAAAAGTTACA 1500
1441 GCACCAATGAGTTGGTGAATGATCGAAATGGGTATTCTTGACCACTAAAGTACA 1500
1501 CATTATGCAATCAAAATCAGCATCAGTAGCTCTTCTTCTTAACTACAGAAATCGCT 1560
1501 CGTTCTGCACTTCAAAATGAGCTTCTATCGCAGGCTTATGATCAACAACAGAGCGATG 1560
1561 ATTGCTGAAAAACCAACCACTTAAAGAGATATGCTTATGCTGCGGCTGATGCTGCT 1620
1561 GTTG-----CAGAGCTTCTAAAGAGACTCTGAGGTTGCTGCTGCTGCTGCTGCT 1611
1621 ATGGGTGTTATGACCGTATG 1641
1612 ATGGGCGCATGGCGCATG 1632

RESULT 2

US-10-329-960-1

; Sequence 1, Application US/10329960

; Publication No. US20030099277A1

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag

; FILE REFERENCE: PBI86PI

; CURRENT APPLICATION NUMBER: US/10/329,960

; PRIOR FILING DATE: 2003-01-02

; PRIOR APPLICATION NUMBER: US 09/643,990

; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: US 08/487,429

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1830121

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4747)..(4747)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (9921)..(9921)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (10150)..(10150)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (29298)..(29298)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (36543)..(36543)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

```

; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Query Match 44.1%; Score 725.8; DB 15; Length 1830121;
Best Local Similarity 65.9%; Pred. No. 2.7e-137; Indels 6; Gaps 2;
Matches 1086; Conservative 0; Mismatches 557

QY 1 ATGGCTCTTAAAGAAATCCCTTTTGTATGCTAAAGCCCGTGAAAACTTTACGAGGTGA 60
DB ATGGCAGCAAAAGACGCTAAATTTGGTAATGACGACGCGTAAATGCTTAAAGCGTG 565409

QY 61 GATAAATCTGCAATCTGTAAAGTAAACACTTGGACCTAAAGCCGTAATGCTGTTAT 120
DB 565410 AATGATTAAGCGGATGACGTAAGTAACCTTTGGCCGGAAGGTCGTATGTAATTTA 565469

QY 121 GAAAAGTCTTTGGTTCCCGAGTTATACAAAAGATGTTGTTGTCGCAAGAAAT 180
DB 565470 GATAAATCAATTTGGCCGCAACAACTACTCAAAAGCGGTGCTGTTGCTCGTGAATC 565529

QY 181 GAACCTGGAATGAAGTTTGAATAATATGGCGCTCAAAATGTTTAAAGAGTAGTCCCAA 240
DB 565530 GAATTAGAAGATAAATTCGAAACATGGCGGCACAAATGTTGAAAGAGTGGCATCTAA 565589

QY 241 ACTAGCGATATTGCTGGTGAAGTAAACAGCAAGTCTTGCACAGCTATTAT 300
DB 565590 GCAAAATGATGCTGCAGTGACGCTACAACTACTGCAACTGTACTTGGCGAAGCTATCGTA 565649

QY 301 CCGTAAGGTGTAAGAACTCTGACGAGTGTCTGTAATCTTATGCGCAATTAAGCGTGCATA 360
DB 565650 AATGAAGGCTTGAAGCAGTACTGCGAGTATGAATCCAAATGATTTAAACGTTGTAAT 565709

QY 361 GATAAGCTGTTGTTGCTGTTTAAAGAACTAAGCGACATTAAGAGCTTACTCGTGAC 420
DB 565710 GATAAGCAGTAAGTCGCGTGTGTTCTGAACCTTAAATTTATCTAAACCTTGTGAACA 565769

QY 421 CAAAGAAATAGCTCAAGTTGGAACCAATTCGMAACTCTGATACAAATAGTAT 480
DB 565770 GCAAAAGAAATTAACAAAGTAGGAGTATTTCTGCAAACTCTGACAGCAATTTGGGTCAA 565829

QY 481 ATCATAGCTGAAGCTATGCTGTAAGTTGAAAGAGGAGTGTATACAGTTGAGGAAGCT 540
DB 565830 TTAATCTCAAGCANTGAAGAAAGTGGCAAGAGGTGTAATTAATCTGTAAGTGT 565889

QY 541 AAAGTCTTGAACATACATAGATGTTGTAAGGAATGAAGTTGACCGTGGCTACCTC 600
DB 565890 ACGGTCTTGAAGATGAATAGATGTTGTTGAAGGATGCAATTCGACCGTGGTACCTT 565949

QY 601 TCTCCATCTTTGTAATCTCTGAGAAATGGTTTGTGAACTTGATACCTTTATATC 660
DB 565950 TCTCCATTTCTCATCAACAAACAGAACTGCAACGTTGTAATAGATAATCCATCTT 566009

QY 661 CTTTGTAAATGAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAGTT 720
DB 566010 CTTCTGTAGATAAAAATCTCAATCTGTAATTAATCTTCCGCTGTAGAGCGTT 566069

QY 721 GCTAAAGTAAACGCTCCCTCTTATTTATGCTGAAGAGTGAAGGTGAAGCACTTGA 780
DB 566070 GCGAAAGCAGGTAACCGTTATTAATCATCGTGAAGAGTGAAGCGGAGCGGTTGA 566129

QY 781 ACACCTGTAGTCAATAAGCTCGTGGAGCACTCCAGTGTAGCCGTAAAGCTCTGCT 840
DB 566130 ACCCTAGTGTAACACTATGCGCGGTATCGTGAAGTTTGCAGCCGTGAAGACCAAGT 566189

QY 841 TTTTGTGAACCGCGTAAAGCTATGCTTGAAGATATTGCTTACTCTTACTGAGGAGAAC 900
DB 566190 TTTTGTGATCGTGAAGCGGATGTTACAGATATTGCAATTTTACACGGGTACAGTG 566249

QY 901 ATATTGTAAGATCGTGGTATAAAGCTTGAAGATGTAAGCTTGTCTTTTAGGAACAGT 960
DB 566250 ATTCTGAAGAAATTTGGTATGAGCTTGAAGAAAGCAACATTGGAAGATTTAGGTCAAGCA 566309

QY 961 AAACGTGTAGTTATTGACAAAGAAATATCTACTATCTGTTGATGGTCTGGAATCAGAA 1020
DB 566310 AAACGTGTGTTATTCAATAAAGATAACACAACTATTGATGGTATCGCGATGAAGCA 566369

QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATGAAAGAAACAGCTCAGATTAT 1080
DB 566370 CAAATCAAGGTGCTGTGCTCAATTCGTGACAAATCGAAGAACTCACTTCTGACTAC 566429

QY 1081 GATCGTGAAGAACTTCAAGAACGCTTTGCAAACTGAAATGAAAGAGAGAGTCTGTTATCCAT 1140
DB 566430 GACAAAGAAAGAACTCCTCAAGACGCGTGTAAATTAGCAGGCGGTGTCAGTAATCAAA 566489

QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGAGTCTGTTAGAGATGCTGTTATCCAT 1200
DB 566490 GTGGCGCAGCACTGAAGTTGAAATGAAAGAGAGAGAGTCTGATGATGATCAATTA 566549

QY 1201 AATGCAACAAAGAGCTGCGGTTGAAAGAGGTATTGTCCTGCTGCTGCTGCTGCTGCT 1260
DB 566550 CATGCAACTCTGTCAGCGGTTGAGAGAGTATCTGTTGCTGCTGCTGCTGCTGCTGCT 566609

QY 1261 CCGTC---CATTAAGTCTTGAATGATATTAACCTGCTGATGATGATGATGATGATGATG 1317
DB 566610 CGCGCASSAGCGAAAGTTGCAGCAAGCTTAAAGAGTGAACAGCAAGCAACAAATGTGGT 566669

QY 1318 CTTAATATCATCGCTGCTTCTTGAAGAGCCTTTACGTCAAATGCTCAATGCTCAATGCTGC 1377
DB 566670 ATCAACTTGCATTCAGTCTATGAGAGCGCTTTACGTCAAATGCTCAATGCTCAATGCT 566729

QY 1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCTGAAACCAAAAGATGGTTTGGATTAAAT 1437
DB 566730 GAAGAGCTTCACTTGTGCAAGTGGGTTAAAAATGGCGAAGAACTTTGGTTATAT 566789

QY 1438 GCTGATCAGGAGATATGAACCTTATTAAGTGGTGTGCTGATGATGATGATGATGATGATG 1497
DB 566790 GCAGAAACAGACAGTACGCGGATATGATCGAAATGGGTATCTTAGATCCAACTAAGTC 566849

QY 1498 ACACGTATTGATTAACAAATCGACATCAGTACGCTCTCTTACTTCTAATCAGATGC 1557
DB 566850 ACTCGTTCTGATTCGATTCGAGCTTCTGAGAGGCTTAAATGATTAATGATTAATGAT 566909

QY 1558 GCTATTGCTGAAAAACCAAGCTTAAAGAGATATGCTGCTGCGGCTGCTGATGGGT 1617
DB 566910 ATGGTAACTGATCTTCCAAAGATGATAAG---CCGATTTAGGTGCTGCTGAAATGGC 566966

QY 1618 GGTATGGGTGATGAGCGGTATGACTA 1646
DB 566967 GGTATGGGTGAAATGGCGGAATGATGA 566995

RESULT 3

US-10-369-493-33982
Sequence 33982, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE OF INVENTION: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR APPLICATION NUMBER: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33982
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Cytophaga hutchinsonii
US-10-369-493-33982

Query Match 43.6%; Score 718.2; DB 12; Length 1626;
Best Local Similarity 65.7%; Pred. No. 5.8e-137;
Matches 1063; Conservative 0; Mismatches 553; Indels 3; Gaps 1;
QY 8 CTAAGAAATCCCTTTTGTAGTCTAAAGCCGCTGAAATACTTTCACGAGGTGTAGATAAAC 67
DB 5 CAAAGCAATTTTATTTGACAGAGACGCCCGCGAATAATTGAAATAAGTGTGTGTCAT 64
QY 68 TTGCAATGCTTTAAAGTAACACTTTGGACCTTAAAGCCCGTAATGTCTGTTATTGAAAGT 127
DB 65 TGGCAGATGAGTAAAGTAAACCTTGGCCCTTAAAGGAAGAAATGTAATTTTAGATAAAA 124
QY 128 CTTTGTGTTCCCGAGTTATTACAAAGATGCTGTATCTGTTGCAAAAGAAATTTGAAGTTG 187
DB 125 AATTCGGTTCACCTGCTATCACAAAGATGGGGTTCTGTAGCAAAAGACATTTGAGTTGA 184
QY 188 AAGATAAGTTTGAATAATATGGGGCTCAAAATGTTAAAGAAAGTAGCTCCCAAAACTAGCG 247
DB 185 AAGATGCTATCGAATAATGCGCGCACAAATTAGTTAAAGAAAGTTGCATCAAAACCGCTG 244
QY 248 ATATTCGTGTGTAAGTAACACTTAACAGCAACAGTCTTGTGCAAGCTATTATTTCGTGAAG 307
DB 245 ATCAGGCTGTGACGGGTACAACTGCGACTGTGTTTAAAGCGAGCAATTTTCAATGCG 304
QY 308 GTCTAAACCTGTAGCAGCTGTCGTAATCCTATGCGCATTAACGTCGGCATAGATAAG 367
DB 305 GTTTAAACAACTGTGAGCGCGCGTAAACCAATGGAATCTTAAAGAGAGGTATTGATAAG 364
QY 368 CTGTTGTGCTGTTACTAAAGAACTAAGCGACATTAACAAAGCTACTCGTGACCAAAAG 427
DB 365 CAGTTTCGGATCAITGAGATCTTAAAGTTTAAAGTTTAAAGAAATCTTAATTCAAAG 424
QY 428 AATAGCTCAAGTTGAAACCATTTCTGCAAACTCTGATAACAATAAGTAAATATCATAG 487
DB 425 AGATGCGACAGTTGCTACAACTCTGCCAACACGACCAAGAAATCGGTAAAGATGTTG 484
QY 488 CTGAAGCTATGCTAAAGTTGAAAGAGAGGTGTTATCAGAGTTGAGAGCAATTAAGTTC 547
DB 485 CACATGCAATGGAATAAGTTGGTAAAGATGCTGTTATTACTGTTGAAGAAAGCTAAAGGTA 544
QY 548 TTGAAACTACATTTAGATGTGTTGAAGGAATGAAGTTTGACCGTGGCTACTCTCCAT 607
DB 545 CTGAACAGAGTAATAAACTGTTGAAGGTATGCAATTTGACCGTGGATATCTATCGCTT 604
QY 608 ACTTTGTAATACTCTGAGAAATGTTTGTGAACCTTGTAATACCCCTTATATCTTTGTA 667
DB 605 ACTTCGTTACGATAACAGATAAATGAAGTTGAGTTAGTAGTCTCCATATATTTCTGATCT 664
QY 668 ATCAGAAAGATTAAGTATGAGATGAAGACATGCTACCAATCTTAGACAAGTTGCTAAAG 727
DB 665 ACATATAAAGTTTCTTCTTAAGGAATTAATCTCGGATCTTGAACAAGTTGTTCAAGT 724
QY 728 TAAACCGTCCACTCTTATTATTGCTGAAGAGCTAGAGGTGAAGCACTTGAACACTTG 787
DB 725 CCGAAGACCGTTGTTAATCATTTGCGGAGATGTTGACGTTGAAGCTATAGCTACACTTG 784
QY 788 TAGTCAATAAGCTCCGTTGAGCACTCCAAGTTGTAGCCGTAAGAGCTCTCTGTTTGGTG 847
DB 785 TTGTAACAATAATCCGTTGGTGAATGAAATCGGGCTGTTAAAGCTCCGGGCTCGGTG 844
QY 848 AAGCCGTAAGCTATGCTTGAAGATATTCCTATCTTACTGAGGAGAGCAATATTG 907
DB 845 ACAGAAAGAAAGCTATGTTGGAAGACATCGCAATTTTAAAGCGGCTGACAGTTATTCTG 904
QY 908 AAGATCGTGTATATAAGCTTTGAAATGTAAGCTTGTCTCTTTTAGGAACAGCTAAAGCTG 967

DB 905 AAGAAAGCGGATACAAATTTAGAAACCGTACACTTGAATATTTAGGTACAGTGAAGA 964
QY 968 TAGTTATTGACAAAGAAATTAATCTATCTGTTGATGTTGCTGGAATAATCAGAGATATTA 1027
DB 965 TCAATATTGATAAAGACAACACACACCGTGTGTTAATGTTTAAAGATTAATATCG 1024
QY 1028 AAGCTGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGATCGTG 1087
DB 1025 TTGACGTTTAAACGATCAAAATCTCAGATGAAATACTACGTCGATTAATCGATCGG 1084
QY 1088 AAAAATCTTCAAGAACGCTTTTGCAAAATCTTGTGTGAGTAGCTGTTTATCCATGTTGAG 1147
DB 1085 AAAAATCTTCAAGAACGCTTTTGCAAAATCTTGTGTGAGTAGCTGTTTATCCATGTTGAG 1144
QY 1148 CTGCTACTGAACTGAATGAAGAAAGAGAGATGTTGTAGAGATGCTCTTAATGCAAA 1207
DB 1145 CAGCTACAGAAAGTTGAAATGAAAGAGAAAGACCGTGTGAGATGATGCAATGCACTA 1204
QY 1208 CAAGAGCTGCGTTGAAGAGGTATTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
DB 1205 CAGCTGCTGCTGTTGAAGAGGTATTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1264
QY 1268 TTAAGTCTTGTGATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1327
DB 1265 GTGCTGCTTGTAGATAAGTTGCAATTCACAAATGAAGATGAAGAAACAGGCAATTCAGATCA 1324
QY 1328 TCGCTGCTGCTTGTGAAGAGCCTTTTACCTCAAAATGCTGCAAAATGCTGCTGCTGCTGCTGCT 1387
DB 1325 TCGCTGCGGCAATTTGAAATCTTCTTACGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
QY 1388 CTATTGTTGAGAAAGTTTCGTGAACCAAAAGATGTTTGTGATTTAAATGCTGCTGCTGCTGCT 1447
DB 1385 CTGTAATGTTGAAAGGTGAAGAGGACAGGCAATTCAGGATCAATGCGCGTGAAG 1444
QY 1448 GAGATATGAAGACCTTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507
DB 1445 ACAGATACGAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1504
QY 1508 CATTACAAATGAGCATGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1567
DB 1505 CGTTAGAAATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1564
QY 1568 AAAACCGAGAACCTTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1623
DB 1565 ACAGCGGAGAAATAATCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1623

RESULT 4

US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 43.1%; Score 710; DB 10; Length 640681;
Best Local Similarity 64.5%; Pred. No. 3.1e-134;

QY 68 TTGCAATGCTGTTAAAGTAACACTTGGACCTAAAGCCGCTAATGTCGTTATTGAAAGT 127
DB |||||
QY 65 TTGCTGACGAGTAAAGTAAGTAACTAGGACCAAAAGGTGTAACGTCGTTCTTTGAAAAGA 124
DB |||||
QY 128 CTTTTCGGTCCCGCAGTTATTACAAAAGATGCTGTATCTGTGCAAAAAGAAATGAACTTG 187
DB |||||
QY 125 AATTCCGTTCTCACTATTACAAATGACGCTGTACCATCGCAAGGAATCGAATTAG 184
DB |||||
QY 188 AAGATAAGTTTGAATAATATGGCGCTCAAAATGGTTAAAGAAAGTAGTCCCAAACTAGCG 247
DB |||||
QY 185 AAGATCGTTTGAATAATGCGTCAAAAGCTTGTGCAAGAGTGGCAAGCAAGCAAAATG 244
DB |||||
QY 248 ATATTGCTGTGATGAGTACACACACACACACCTCTTGCACAGCTATTATTTCGTTGAAG 307
DB |||||
QY 245 ACATTGCTGGGACGGTACACACACACACACGCTCTTGCACAGCGATGATTTCGGGAAG 304
DB |||||
QY 308 GTCTAAACCTGTAGACGCTGGTGGTAAATCCTATGGCCATTAAACGCTGGCATAGATAAG 367
DB |||||
QY 305 GCTTGAATAACGTCACGCTCTGGCGGNAACCAATGGTGAATCGTAAAGGAATCGAAAAG 364
DB |||||
QY 368 CTGTTGTTGCTGTACTAAAGAACTAAGCGACATTAACAAGCCTACTCGGTGACCAAAAAG 427
DB |||||
QY 365 CGACTCAAGTAGCGGTTGAAGAACTTCTTAAATCTCTAAGCCAAATCGAAGGCAAGATT 424
DB |||||
QY 428 AATAGCTCAAGTTGNAACCACTTCTGCAACTCTGTATACAACAATAGGTAAATCATAG 487
DB |||||
QY 425 CAATCGCTCAAGTTGCGCGCATTTCTTC---TGTGACGATGAAGTAGGGAATAATCATG 481
DB |||||
QY 488 CTGAAGCTATGGCTTAAAGTTTGGAAAAGGAGTGTATCACAGTTGAGGAAGCTTAAAGGTC 547
DB |||||
QY 482 CTGAAGCAATGGAGCGGTAGGCAACGACGCGGTATTACGATTGAAGAGCTTAAAGGAT 541
DB |||||
QY 548 TTGAACACTACATTAGATGCTGTGAAGAAATGAAGTTTGACCGGTGCTACTCTCCAT 607
DB |||||
QY 542 TCTCTACAGAGCTAGAAAGTGGTTGAAGTATGCAATTTGACCGCGCTATGCGCTCTCCT 601
DB |||||
QY 608 ACTTTGTAACCTAATCTCGAATAATGTTTGTGAACCTTGATTAACCTTATCTTTGTA 667
DB |||||
QY 602 ACATGGTGAAGAGCTCTGCAAGATGAAGCGGTTCTTGACACCCCTTATGTCTTGATCA 661
DB |||||
QY 668 ATGAGAAAAGATTACTAGCATGAACACATGCTTACCAATCTTGAACAAGTTGCTTAAAG 727
DB |||||
QY 662 CAGATAAGAGATTCTTAAATCCAGAGATTAATACCAGTTCTTGACCAAGTTGTACAAC 721
DB |||||
QY 728 TAAACCGTCCACTCTTATTATTGCTGAAGACGTPAGAAGTGAAGCACTTGAACACTTG 787
DB |||||
QY 722 AAGCAAGCCAAATCCTTATCCTGCTGAGGATGTAGAAGGTGAAGCTCTTGCAACACTTG 781
DB |||||
QY 788 TAGTCAATAAGCTCCGTTGAGCACTCCAAAGTTGTAGCCGTAAAAGCTCTGTTTGGTG 847
DB |||||
QY 782 TGGTGAACAAACTTCGTGGGAACATTTAACCGAGTGAAGGTTTAAAGCTCTCGGATCGGT 841
DB |||||
QY 848 AACCCGTTAAAGCTATCTGTTGAAGATATTGCTATCTCTTACTGAGAGAGAAGCAATATTG 907
DB |||||
QY 842 ATCGTGTAAAGCAATGCTTGAAGCAATGCGATCTTACTGCGGTGAAGTATCACAG 901
DB |||||
QY 908 AAGATCGTGTATTAAGCTTGAATAATGTAACTTGTCTTTTAGGAACAGCTTAAACGTG 967
DB |||||
QY 902 AAGACCTTAGGCTTGACTTGAATCTCGAATCATCACACAGCTTGGTCGCGCAAGTAAAG 961
DB |||||
QY 968 TAGTTATTGACAAAATACTACTATCTGTGATGCTGGAATCAGAAATCAGAAATATTA 1027
DB |||||
QY 962 TCGTTGTAACGAAGAGAACAAACGATTTGTTGAAGGTGCTGGCGAAGCGATAAATCG 1021
DB |||||
QY 1028 AAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATCATCGT 1087
DB |||||
QY 1022 CAGCTCGGTTTAAATCAAAATCAAGCAAAATTCGAAGAAACAAACATCTGACTTCGATAAG 1081
DB |||||
QY 1088 AAAAAGCTTCAAGAACGCTCTTGCAAACTTGTGTTGGTAGTAGCTGTTTATCCATGTTGAG 1147
DB |||||
QY 1082 AAAAAGCTTCAAGAGCGCTTAGGAAGCTTGTGTTGGCGCTCGCTGTTCTTAAAGTCGCTG 1141
DB |||||

QY 1148 CTGCTACTGAACTGAAATGAAGAGAGAAGGATCGTGTAGAAGATGCTCTTAAATGCAA 1207
DB |||||
QY 1142 CAGCTACTGAACTGAAATGAAGAGCGCAAACTTCGATTTGAAGCGCTTGAAGCTCTA 1201
DB |||||
QY 1208 CAAGAGCTCGGTTGAAGAGGATTGTCCTCGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
DB |||||
QY 1202 CTCGCGCGCTGTAGAAGAGGAATCGTTCTGCTGGTGGTACTGCTCCCTTGTGAACCTGA 1261
DB |||||
QY 1268 TTAAGTCTCTGATGATATTAAACCTGCTGATGATGAACCTTGTGGACTTAAATATCA 1327
DB |||||
QY 1262 TTAAG---CAGTCTCTAGCATCGGTGCAAGAGGTGACGAAGCAACAGGTGTGAACATCG 1318
DB |||||
QY 1328 TCCTGCTCTCTCTTGAAGAGCCTTTACCTCAAAATGCTGCAAAATGCTGGCTATGAAGGTT 1387
DB |||||
QY 1319 TCTTCTGCTGCTTGAAGAGCAGTTTCGTCAAATTCCTCAACAGCAGGCTTTGAAGGCT 1378
DB |||||
QY 1388 CTATTGTTGTAGAAAAGTTCTGTAACCAAAAGATGTTTTGGATTAAATGCTGCTATCAG 1447
DB |||||
QY 1379 CTGTGATCGTCGAGCGCTCAAGAAAAGAGAGCTGCTTCGGTTTCAACGCGCAACTG 1438
DB |||||
QY 1448 GAGAATATGAAGACCTTATTAAAGCTGCTGCTCAATTCCTTAAATAAGTTTACAGTATTG 1507
DB |||||
QY 1439 GTCAATGGGTGAACATGTTGAAGCTGGTATCGTTGACCCCAAGCAAAAGTAACTCGTTCAG 1498
DB |||||
QY 1508 CATTACAAAATGACAGCATCAGTAGCCTCTTACTTCTTAACTACAGAAATGCGCTATTTCGTG 1567
DB |||||
QY 1499 CGCTTACGACGCGAGCATCTGTATCCGCTATGTTCTTCAACAGAGCGGTGATCGCTG 1558
DB |||||
QY 1568 AAAAAACAGAACCTTAAAAAGATATGCTATGCTGCGGTGGTATGCGGTGTTGCGGTG 1627
DB |||||
QY 1559 ATAAGCCTGA---GGAGAACGAGCGCGCGGCGGAATGCTGACATGGCGGAATGCGGTG 1615
DB |||||
QY 1628 GTATGACGCTATCTACTA 1646
DB |||||
QY 1616 GCATGGCGCGCATGATGTA 1634
DB |||||

RESULT 6

US-09-070-927A-42
; Sequence 42, Application US/09070927A
; Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch
Steven Barash
Patrick J. Dillon

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

Db 177912 GCCTTATCGCTGATATCCAGAGAGAA 177883

RESULT 8
US-10-267-311-50
; Sequence 50, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267,311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613,303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1923)
US-10-267-311-50

Query Match 41.9%; Score 689.8; DB 15; Length 1926;
Best Local Similarity 65.8%; Pred. No. 4e-131;
Matches 1035; Conservative 0; Mismatches 532; Indels 6; Gaps 2;

Qy 8 CTAAGAAATCCCTTTGATGCTAAAGCCCGTGAAAACTTTCACGAGGTGTAGATAAC 67
Db 5 CAAAGAAATTAATTTTTCATCAGATCCCGTTCAGCTATGGTCCGGTGTGATATCC 64

Qy 68 TTGCAATGCTTTAAAGTAACACTTGGACCTAAAGCCGTAATGTCGTTATTCAAAAGT 127
Db 65 TTGCAATGCTTTAAAGTAACACTTGGACCTAAAGCCGTAATGTCGTTATTCAAAAGT 124

Qy 128 CTTTGTGTTCCCAAGTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATGAAGT 187
Db 125 CATTCGGTTCACCTTGATTACCAATGACCGTGTGACTATTGCCAAAGAAATGAATTAG 184

Qy 188 AAGATAAGTTTGAATAATGGCGCTCAATGTTAAAGAGTAGTCCCAAACTAGCG 247
Db 185 AAGACCAATTTGAAATATGGGTGCCAAATTTGGTATCAGAGTAGTCTTCAAAACCAATG 244

Qy 248 ATATGCTGTGTAGGAAGTACAAAGCAAGTCCCTTGGCAAGCTATTATCGTGAAG 307
Db 245 ATATCGAGGTGATGGAAGTACAAAGCAAGTCCCTTGGCAAGCTATTATCGTGAAG 304

Qy 308 GTGTAAAGAACTTGTAGCAGCTGGTGTATATCCATCGTATGGCCATTAACGTTGGCATAGATAAG 367
Db 305 GAATCAAAAAGCTCACAGCAGGTGCAAAATCCAAATCGGTATTTCGTCGTGGGATTGAAAACAG 364

Qy 368 CTGTTGTGCTTACTTAAGAACTAAGCAGCATTAACAGCCCTACTCGTGACCAAAAG 427
Db 365 CAGTTGGCGCAGCAGTGTGAAGCTTTGAAAAACAAGCTATCCCTGTTGGCCAAATAAGAG 424

Qy 428 AAATAGCTCAAGTTGGAGCACTTTCTGCAAACTCTGTATACAACTAGTAAATATCATAG 487
Db 425 CTATGCTCAAGTTGCAGCCGTATCTTCTCGTTCTGA---AAAAGTTGGTCAAGTACATCT 481

Qy 488 CTGAAGCTTATGGCTTAAAGTTGGAAGAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db 482 CTGAAGCAATGMAAAAGTTGCAAAAGACCGTGTGTCATCACCATCGAAGAGTCAAGTGGTA 541

Qy 548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db 542 TGGAAACAGAGCTTGAAGTCGTAGAGAGANTGAGTTTGACCGTGGTACCTTTTCACAGT 601

Qy 608 ACTTTGTAACATACTCTGAGAAAAATGGTTTGTGAACCTTGATTAACCTTTATTCCTTTGTA 667
Db 602 ACATGGTGCAGATAGCGAAAAATGGTGGCTGACCTTGAAAAATCCGTACATTTTGATTA 661

Qy 668 ATGAGAAAAGATTAAGCATGAAGACATGCTACCAATCTTAGAACAAAGTTGCTAAAG 727
Db 662 CAGACAAGAAAAATTTCCAAATATCCAGAAATCTTGCCACTTTTGGAAAGCAATTCCTCAA 721

Qy 728 TAAACCGTCCACTCTTTATTATTCTGAAGACGTAGAGGTGAAGCACTTTGCAACACTTG 787
Db 722 GCATCGTCCACTCTTTGATTTATTCGGATGATGTGGATGTGAGGCTCTTCCAACTCTTG 781

Qy 788 TAGTCAATAAGCTCCGTTGGAGCACTCCAAAGTTGTAGCCGTAAGAAAGTCTCTGGTTTGGTG 847
Db 782 TTTTGAACAAGATTCGTGGAACCTTCAACGTAGTAGCAGTCAAGGCACCTGGTTTGGTG 841

Qy 848 AAGCCGTAAAGCTATGCTTGAAGATATTGCTATCTTACTTGGAGGAGAGCAATATTG 907
Db 842 ACCGTGCAAAAGCCATGCTTTGAAGATATCCCATCTTAAACGCGAACAGTTATCAAG 901

Qy 908 AAGATCGTGTATATAAGCTTTGAAAAATGAAGCTTGTCTTTTAGGAACAGCTAAACGTG 967
Db 902 AAGACCTTGTCTTGTGTTGAAGATGCGACAATTTGAAGCTCTTTGGTCAAGCAGCGAGAG 961

Qy 968 TAGTTATTGCAAAAGAAATPACTATCTTGTATGTTGTGCTGTAAGAAATCAGAGATATTA 1027
Db 962 TGACCGTGGACAAAGATAGCACGGTTATTGTAGAAGGTGACGAAATCTCTGAAAGCGATT 1021

Qy 1028 AAGCTGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGATCGTG 1087
Db 1022 CTCACCGTGTGGGTATCAAGTCTCAAAATCGAAACTACAACTCTCGAATTTGACCGGTG 1081

Qy 1088 AAAAATCTCAAGAACGCTTTGCAAACTCTGTGTGTGAGTAGCTGTTATCCATCTTGGAG 1147
Db 1082 AAAAATGCAAGAACGCTTTGCCAAATTTGTCAAGTGTGTAGCGTTATTAAAGTTCGGAG 1141

Qy 1148 CTGCTACTGAAACTGAAATGAAAGAGAAAGGATCGTGTAGAAGATGCTCTTAAATGCAA 1207
Db 1142 CCGCAACTGAAACTGAGTTGAAAGAAATGAAACTCCGCAATTTGAAGATGCCCTCAACGCTA 1201

Qy 1208 CAAGAGCTCGGTGAAAGAGTATTCTCCCTGGTGGTGTACTGCTTTTGTCCGCTCCA 1267
Db 1202 CTCGTGAGCTGTTGAAGAGGTTATTGTGAGGTGGTGAACAGCTCTTTCCTCAATGTGA 1261

Qy 1268 TTAAGTCTTGTATGATATTAAACCTCTGTATGATGATGAATTCCTGGACTTAATATCA 1327
Db 1262 TTCAGCTGTTGTACCTTGGAAATTGACAGAGATGAAGCA--ACAGGAGTAAATTG 1318

Qy 1328 TCCGTGCTTCTTTGAAGAGCCTTTACGTCAAAATGTGTGCAAAATGCTGGGTATGAAGGTT 1387
Db 1319 TTTCCGCTGCTTTGGAAGAACCTGTTCTGTCAAATTTGCTCAATTCAGAGATTTGAAGGAT 1378

Qy 1388 CTATTGTTGAAAAAGTTCGTGAACCAAAAGATGGTTTGGATTTTAAATGCTGCTCATCAG 1447
Db 1379 CTATCGTTATCGATCGTTTGAAGAAATCTGAGCTTGTGTATAGGATTAACCGCAACCTG 1438

Qy 1448 GAGAAATGAAGACCTTTATTAAAGCTGTGTGATTTGATTCCTTAAAAAAAGTTTACAGTATTG 1507
Db 1439 GCGAGTGGTTAATCATGATTTGATCAAGGTATCATTTGATCCAGTTAAAGTGAAGTCTTCAG 1498

Qy 1508 CATTACAAAATGACAGCATCAGTACGCTCTTACTTCTTAACTTACAGAAATGCGCTATTGCTG 1567
Db 1499 CCCTACAAAATGACAGCATCTGTAGCCAGCTTGATTTTGAACAACAGAGCAGTCTGTGCA 1558

Qy 1568 AAAAACAGAAC 1580
Db 1559 ATAACCAGAAC 1571

RESULT 9

US-09-841-132-380
; Sequence 380, Application US/09841132
; Patent No. US20020061848A1

GENERAL INFORMATION:

; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 380
; LENGTH: 1635

TYPE: DNA

ORGANISM: Chlamydia pneumoniae

US-09-841-132-380

Query Match 41.9%; Score 689.6; DB 9; Length 1635;

Best Local Similarity 65.1%; Pred. No. 4.1e-131;

Matches 1035; Conservative 0; Mismatches 549; Indels 6; Gaps 1;

QY	1	ATGGCTTCTAAAGAAATCCTTTTGGATGCTAAAGCCCGTGAAGAACTTTTCCAGAGGTGA	60
DB	1	ATGGCAGCGAAATAATTAATATATATGAAGAGCCAGAAATAATACATAAGGGTGA	60
QY	61	GATAAATTGCAATCTGTTTAAAGTAACTTTGACCTTAAAGGCGGTAAATTCGCTTAT	120
DB	61	AAAACTTTGCAAGACGAGTAAAGTTACTCTAGGTCTTAAAGGAGCTCAGCTAGTTATA	120
QY	121	GAAAGTCTTTGGTTCCTCCAGTATTAACAAGATGGTGTATCTGTTCAGAAAGAAAT	180
DB	121	GATAAGAGCTTTGGTCTCTCCCAAGTACTAAAGATGGTGTACTGTAGCTAAAGAAATC	180
QY	181	GAACTTGAAGATAAGTTTGAAATATGGCGCTCAAAATGGTTAAAGAGTAGTCCCAA	240
DB	181	GAGCTCGAAGACAACATCAAAACATGGCGCTCAGATGGTAAAGAGTCCCGACAAA	240
QY	241	ACTAGGATATTTGGTGGTATGGAATCAACAGCAACAGTCCCTTGCAACAGCTATTAT	300
DB	241	ACTGCTGACAAAGCAGCGGACGGAATCAACAGCAACAGTCTTCTGCAAGCAATCTAT	300
QY	301	CTGAGAGTGAATCTGTACAGCTGGTCTGATCTTATCTGATCCATTAAGCTGACATA	360
DB	301	ACGAGAGGTCTAAGAAATGTCTACTGCGGTGCAATCTCTATGGACCTAAAGAGGTATC	360
QY	361	GATAAGCTGTGTGTCTTACTTAAAGAACTAAGCGCATTAAGAGCTTACTCGTGAC	420
DB	361	GACAAGCGTAAAGTTGTTGTTGATGACTCAAAATAATAGTAAACCTGTACAACAT	420
QY	421	CAAAAGAAATAGCTCAAGTGGAAACCAATTTCTGCAAACTCTGTATACAAATAGGTAT	480
DB	421	CACAAGAAATCGCTCAAGTAGTACTATCTCAGCAATAATGATTCGGAATCGGAAT	480
QY	481	ATCATAGCTGAAGCTATGCTTAAAGTGAAGAGGAGGTGTTATCACAGTTGAGAGCT	540
DB	481	CTTATGAGAGCTATGGAAGAAAGTTGGTAAAGACCGGATCAATCTTCTGGAAGCT	540
QY	541	AAAGCTTTGAAACTACATTTAGATGGTGTGAAGAAATGAAGTTTGACCGTGGCTACCTC	600
DB	541	AAAGCTTCGAAGCTTCTCAGCTGTGAGAGGAATGAACCTTCAACCGTGGATACCTC	600
QY	601	TCTCCTACTTTGTATCAATCTCTGAGAAATGGTTGTGAATCTGTATGAACCTTATATC	660
DB	601	TCCAGCTACTTCTCCACAATCCAGAACTCAAGAAATGGTTTGAAGACGCTCTGATT	660
QY	661	CTTTGTATGAGAAAGATTTACTAGCATGAAGACATGCTACCAATCTTACAAAGATT	720
DB	661	CTAATCTACGATAAAAAAATCTCTGGAATTAAGACTTCTCTCCAGTTTACAAAGTGA	720

QY	721	GCTAAAGTAAACCGTCCACTCTTATTTATTTGCTGAAGACCTAGAAGGTGAAGCACTTGA	780
DB	721	GCAGAACTCGAGCGCCCTCTTTTATCATTTGAGAGAAATTTGAAGGAGGCTTTAGCA	780
QY	781	ACATTTAGTCAATAAGCTCCGTGAGACACTTCAAGTTGTAGCCGTAAAGCTCTCTGT	840
DB	781	ACTCTAGTCAATAGACTCCGTGAGGATTCAGAGTCTGTGCAAGTGAAGCTCTCTGT	840
QY	841	TTTGGTGAACCGCGTAAAGCTATGCTTGAAGATTTGCTTATCTTACTGAGGAGCA	900
DB	841	TTGCGTGAAGAGAAAGCTATGTTAGAAAGCATCGCTTATCTTACTGCGCCAACTA	900
QY	901	ATATTTGAAGATCGTGGTATAAAGCTTGAAGATTTGAAGCTTCTTCTTTAGGAACAGCT	960
DB	901	GTTAGGAGAACTTTGGCATGAACTAGAGATACAACTCTAGCAATGTTAGGAAGCT	960
QY	961	AAACGTGTAGTTATTGACAAAGAAATACTATCTGTTGATGGTGTCTGAAATCAGAA	1020
DB	961	AAGAAAGTTATCGTAACTAAAGAGATACCAATCTGTCGAAGGCTTAGGAAACAACT	1020
QY	1021	GATATTAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGACTCAGATTAT	1080
DB	1021	GATATCCAGCTCGATGCGCAATATTTAAAGAACTTGAAGATAGCACTTCAAGATTAC	1080
QY	1081	GATCGTGAAGAACTTTCAAGAACGCTTTGCAAACTTTGTTGGTGGAGTAGCTGTATCCAT	1140
DB	1081	GACAAAGAAATCTCCAGAGCGTTTGTAACTCTCCGTTGGTGTCCCGTAACTCCGC	1140
QY	1141	GTGGAGCTGTACTGAAACTGAAATGAAAGAGAGAGGATCGGTGAGAAATGCTCTA	1200
DB	1141	GTAGAGCTGTCTACCGAAATAGAGATGAAAGAGAGAGAGAGATGATGATGACCAA	1200
QY	1201	ANTGCAAGAGCTGCGGTTGAAAGAGTATTTGCTCCCTGCTGGTGGTACTGCTTTGTC	1260
DB	1201	CAGCAACCATTTGAGCTGTGCAAGAGGAAATCTCCCTGGTGGTGGTACTGCTTTGTT	1260
QY	1261	CGCTCCATTAAGCTCTTGTATGATATTAACCT-----GCTGATGATGATGAACTTGCT	1314
DB	1261	CGCTGTATCCCTACACTAGAACCTTTCTTCTATGCTAGCAACGAGAGCAAGCTATT	1320
QY	1315	GGACTTAATATCATCGCTGGTCTCTTGAAGAGCTTTAGCTCAATTTGCTGCAATGCT	1374
DB	1321	GGTACTCGTATTTTCTAAAGCAATTAACAGCTCCATTAAGCAAAATTCGAAGTACGCA	1380
QY	1375	GGCTATGAGGTTCTTATTTGCTAGAAAGGTTCTGTAACCAAGATGGTTTGGATT	1434
DB	1381	GGTAAAGAGGCGCTATCATTTGTGAGCAAGTTCTAGCAAGATCTGCAAAATGAAGCTAT	1440
QY	1435	AATGCTGATCAGGAGATATGAAGACCTTTATTAAGCTGGTGTCTATTTGATCCTAAAAA	1494
DB	1441	GATGCTTTACGTGACGCTTATACATATGATTGACGAGGAATTTTAGATCCAATAA	1500
QY	1495	GTTACAGTATTTGCAATTAAGAAATGAGATCAGTACGCTCTTACTTCTTAAGTACAGAA	1554
DB	1501	GTGACTCGCTCAGCTCTAGAAAGCGAGCTTCTATCGAGGATTTACTCTCTCAACAGAA	1560
QY	1555	TGCGCTATTTGCTGAAACCAAGCACTTAA 1584	
DB	1561	GCCTTAATCGCTGATATCCAGAGAGAAA 1590	

RESULT 10

US-09-960-428-13

; Sequence 13, Application US/09960428

; Patent No. US20020115147A1

GENERAL INFORMATION:

; APPLICANT: Roche Diagnostics GmbH

; TITLE OF INVENTION: Method for producing an active heterodimeric AMV-RT in prokaryot

; FILE REFERENCE: 5272/00/

; CURRENT APPLICATION NUMBER: US/09/960,428

; CURRENT FILING DATE: 2001-09-21

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 13
LENGTH: 2155
TYPE: DNA
ORGANISM: Escherichia coli
US-09-960-428-13

Query Match 41.6%; Score 684.4; DB 10; Length 2155;
Best Local Similarity 63.5%; Pred. No. 5.3e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

QY 1 ATGGCTTCTAAAGAAATCCCTTTTGTATGCTAAAGCCCGTGAAGAAATCTTCCAGAGTGTA 60
Db 472 ATGGCAGCTAAAGAGGFAAAATTCGGTAAACGAGCTCGTGTGAAATATGTCGCGCGGTA 531

QY 61 GATAAATCTGCAATGCTGTAAAGTAACACTTGGACCTAAAGCCGTAATGCTTAT 120
Db 532 AAGCTACTGGCAGATGAGTAAAGTATACCTCGGTCCAAAGCCGTAAGCTAGTTCTG 591

QY 121 GAAAAGTCTTTTGGTTCCCGCAGTTATTACAAAAGATGGTGTATCTGTTCGAAAAGAAAT 180
Db 592 GATAAATCTTTCCGTGACCGACCACTACCAAGATGGTGTTCCTGCTCGTGAATC 651

QY 181 GAATTTCAAGATAAGTTTGAATAATATGGCGCTCAATGGTTTAAAGAGTAGTCCCAA 240
Db 652 GAATCGAAGACAAGTTTGAATAATATGGTGGCGAGATGGTGAAGAGTTGCTCTAAA 711

QY 241 ACTAGCGATATTCGTGTGATGTAACACTAACACAGACAGTCTTGCACAGCTATTTAT 300
Db 712 GCAACGACCTGCGAGCGAGCTTACCACTGCAACCGTACTGGCTCAGGCTATCATC 771

QY 301 CGTGAAGGTGTAAACCTTTGACAGCTGGTGGTAACTCTATGGCCATTAAACCTGGCATA 360
Db 772 ACTGAAGTCTGAAGCTGTGTGCGGGCATGAAACCGATGACCTGAAACGTTGATC 831

QY 361 GATAAAGCTTGTGTGTTTACTTAAAGAACTAAGCGACATTAACAAAGCCCTACTCGTGAC 420
Db 832 GACAAAGCGTTTACCGTGCAGTTGAAGAACTGAAGCGGTGTCGTTACCATGCTCTGAC 891

QY 421 CAAAAGAAATAGCTTAAGTTGAAACCAATTTCTGAAACTCTGATACAACTAGGTAAT 480
Db 892 TCTAAAGCGATGTCTCAGTTGGTGTACCATCTCCGCTAACTCCGACGAAACCGTAGTAAA 951

QY 481 ATCATAGCTCAAGCTATGGCTTAAAGTTGAAAGAGGAGTGTATACAGTTGAGGAAGCT 540
Db 952 CTGATCCTGAAGCGATGGAAGAACTCGTAAAGAGGCGTATACCCGTTGAAGCGT 1011

QY 541 AAAGGCTTTGAAACTACATTTAGATGTGGTTGAAGGAATGAAGTTGACCGTACCTC 600
Db 1012 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTTCGACCGTGGCTACCTG 1071

QY 601 TCTCCATCTTTGTAATCTCTGAGAAATGTTTGTGAACCTTGATACCTTATATC 660
Db 1072 TCTCCTTACTTCTCAACAGCGGAACTGGCGCAGTAGAAGCTGGAAGCCCGTTTCATC 1131

QY 661 CTTTGTAAATGAGAAAGATTTACTAGCATGAAGACATGCTACCAATCTTTAGAACAGATT 720
Db 1132 CTGCTGCTGACAGAAATCTCAACATCGCGAATGTCGCGTTCGAAAGCTGTT 1191

QY 721 GCTAAAGTAACCGTCCACTCTTATTATTGCTGAAGACGCTAGAAGGTGAAGCACTTGCA 780
Db 1192 GCCAAAGCAGCAAAACCGCTGCTGATCATCGCTGAAGATGTAAGGCGAAGCGCTGCA 1251

QY 781 ACATTTGATCAATAAGCTCCGTTGGAGCACTCCAAAGTTGAGCCGTAAAGCTCTCGGT 840
Db 1252 ACTCTGTTGTTTAAACCATCGTGGCATCGTGAAGAGTGGTGGTGAAGCAACCGGGC 1311

QY 841 TTTGGTGAACCGCTTAAGCTATCTTTGAAGATATGCTATCTCTTACTGAGGAGAGCA 900
Db 1312 TTCGGCATCTGCTAAAGCTATGCTGAGATATCGAACCCCTGACTGGCGGTACCGTG 1371

QY 901 ATATTTGAAGATCTGGTGTATAAGCTTTGAAATGTAAAGCTTGTCTTCTTTAGGAACAGCT 960
Db 1372 ATCTGAGAGATCGGTATGGACTGGAAGAAACCAACCTTGGAGAGCTGGGTGAGCT 1431

RESULT 11

US-10-157-317-26
; Sequence 26, Application US/10157317
; Publication No. US2003023235A1
; GENERAL INFORMATION:
; APPLICANT: Surber, Mark W.
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Neil Berkley
; TITLE OF INVENTION: MINICELL-BASED SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: AND PROTEINS THAT MODULATE THE ACTIVITY OF SIGNALLING
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: MPX.008DV21
; CURRENT APPLICATION NUMBER: US/10/157,317
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,951
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 26

; LENGTH: 3840
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion protein
US-10-157-317-26

Query Match 41.6%; Score 684.4; DB 12; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

QY 1 ATGCGTCTTAAAGAAATCCCTTTTGATGCTAAAGCCCGTGAAGAACTTTTACAGAGTGTA 60
Db 2133 ATGCGAGCTAAAGACGCTAAATTCGGTAAACGAGCTCGTGTGAAATATGTCGCGCGTA 2192

QY 61 GATAAATCTGCAATGCTGTTAAAGTAACACTTGACCTAAAGCCGTAATGCTGTTAT 120
Db 2193 AACGTACTGGCAGATGCGATGAAGTTACCTCGTCCAAAGGCGGTACGTAGTTCTG 2252

QY 121 GAAAGTCTTTTGGTTCCCGAGTTATACAAAGATGGTGTATCTGTTGCAAGAAAT 180
Db 2253 GATAAATCTTTCGGTGCACCGACCATCACCAAGATGGTGTTCGTTGCTCGTGAATC 2312

QY 181 GAACTGAAGATPAAGTTGAAATATGGCGCTCAAAATGGTTAAAGAAAGTAGTCCCAAA 240
Db 2313 GAACTGAAGACAAAGTTGAAATATGGTGGCGAGATGGTGAAGAAAGTTGCTCTAAA 2372

QY 241 ACTAGCGATATTGCTGGTGAAGTAACACTCAACAGCAAGTCTTGCACAAAGCTATTAT 300
Db 2373 GAAACAGAGCTGCGAGGAGCGGTACCACTGCAACCGTATCGCTCAGGCTATCATC 2432

QY 301 CGTGAAGGTGTAAGACTTTAGCAGCTGGTCTGTAATCTCTATGGCCATTAAGCGTGCA 360
Db 2433 ACTGAAGGTCTGAAGCTGTTCTCGGCGCATGACCCGATGACCTGAAGCGTGTATC 2492

QY 361 GATAAGCTGTTGCTGTTACTAAAGAACTAAGCGACATTAAGAGCTTACTCGTGAC 420
Db 2493 GACAAAGCGTTACCGCTGCAGTGAAGAACTGAAAGCGCTGCTCGTACCATGCTGAC 2552

QY 421 CAAAGAAATAGCTCAAGTTCGAACCATTTCTGAAACTCTGATACACAAATAGGTAT 480
Db 2553 TCTAAGCGATTGCTCAGTTCGTACCATCTCCGTTAATCTCGACGAAACCGTAGGTAA 2612

QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGAAAGAGGAGTGTATCACAGTTGAGAAAGCT 540
Db 2613 CTGATCGCTGAAGCGATGACAAAGTCTGTAAGAGGCGTTATCACCGTTGAAGCGT 2672

QY 541 AAAGTCTTGAACCTACATTAAGTGTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
Db 2673 ACCGGTCTGACGAGCAACTGGACGCTGTTGAAGGTATGCACTTCGACCGTGGCTACCTG 2732

QY 601 TCTCCATCTTTGTAACCTAATCTCGAAGAAATGGTTTGTGAACCTGATACCTTTATATC 660
Db 2733 TCTCCTTACTTCATCAACGCGGAACTGCGCGAGTAGAAGTGAAGCCCGTTTATC 2792

QY 661 CTTTGTAAAGAAAGATTAAGTGAAGCAATGCTTACCAATCTTAGAACAAGTT 720
Db 2793 CTGCTGGCTGACAGAAATCTCAACATCCGCGAAATGCTCGCGTCTTGAAGCTGTT 2852

QY 721 GCTAAGTAAACCGTCCACTCTTATTAATGCTGAAGAGTGAAGGTAAGCACTTGA 780
Db 2853 GCCAAGCAGGCAACCGCTGCTGATCATCGTGAAGATGTAAGAGCGAAGCGTGGCA 2912

QY 781 ACATGTTAGTCAATAAGCTCGTGGAGCACTCCAGTTGTAGCCGTAAAGCTCTCTGT 840
Db 2913 ACTCTGTTGTTAAACACCATGCGTGGCATCTGTAAGTCTGCGGTAAAGCACCAGG 2972

QY 841 TTTGTTGAAGCGGTAAGCTATGCTGAAGATTTGCTATCTTACTCGAGGAGCAAGCA 900
Db 2973 TTCGGCATCGTGAAGCTATGCTGAGGATATCGCAACCTGACTGCGGTACCGTG 3032

QY 901 ATATTGAAGATCGTGGTATAAGCTTGAAGTGAAGCTTCTTTTAGAACAGCT 960
Db 3032

Db 3033 ATCTCTGAAGAGATCGGTATGAGCTGGAAGAAAGCAACCTCGAAGACCTGGGTAGGCT 3092

QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCTGTTGATGCTCGAAATCAGAA 1020

Db 3093 AAACGTGTGTGATCAACAAAGACACCCACTATCATCATGATGGCGTGGGTGAAGCT 3152

QY 1021 GATATTAAAGCTTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACCAAGCTCAGATTAT 1080

Db 3153 GCAATCCAGGCGGTGTTGCTCAGATCCGTACAGAGATGAAGAGCAACTTCTGACTAC 3212

QY 1081 GATCGTGAAGAACTTCAAGAACGCTTTCGAAACTTGTGGTGGAGTAGCTGTATCCAT 1140

Db 3213 GACCGTGAAGAACTCGAGAACGCTAGCGAACTGGCAGGCGGCTTCAGTTATCAAA 3272

QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAAGAAAGAGATCGTGTAGAAAGTCTCTA 1200

Db 3273 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAAAGAAAGACGCGTTGAAGATGCCCTG 3332

QY 1201 AATGCAACAAAGAGCTCGGTTCAAGAAAGTATTGTCCTCGTGGTGGTACTGCTTTGTC 1260

Db 3333 CACGCGACCGCTGCTGCGGTAGAAGAGCGGTGTTGCTGGTGGTGTGCTGCTGATC 3392

QY 1261 CGTCCATTAAGTCTTGTATGATATTAACCTGCTGATGATGAAGTCTGCTGACTT 1320

Db 3393 CGCGTAGCTCTAAACTGGCTGACCTGCGTGTGCTGCAACCAAGACCAAGACCGTGGTATC 3452

QY 1321 ATATCATCCGCTGCTCTTCAAGAGCCCTTTACGTCAAATTCGCTCAAAATGCTGCAATGCTGCTAT 1380

Db 3453 AAAGTTGCACTCGGTGCAATGGAAGCTCGCTGCGTCAGTCTGATTTGAATCTGCGCGAA 3512

QY 1381 GAAGGTTCTATTGTTGTAGAAAGTCTGTGAACCAAGATGGTTTTGGATTTATGCT 1440

Db 3513 GAAACGCTCTGTTGTTGCTTAACACCGTTAAAGCGCGCGCAACTACGTTTACACGCA 3572

QY 1441 GCATCAGAGAAATATGAAGACCTTATTAAGCTGTGTGCTCATTTGATCCCTPAAAAAGTTACA 1500

Db 3573 GCAACGGAAGATACGCAACATGATCGACATGGGTATCTGTGATCCCAACCAAGTAACT 3632

QY 1501 CGTATTGCAATCAAAATGCAGCATCAGTAGCTCTTACTTCTTAACATCAGAAATCGCT 1560

Db 3633 CTTCTGCTCTCAGTAGCAGCTCTTGTGGCTGGCTGATGATCACCAAGTATGATG 3692

QY 1561 ATTGCTGAAAAACCAAGACCTAAAAAGATATGCTCTATGCTGGCGGTGCTATGGTGGT 1620

Db 3693 GTTACCGACCTCGCGAAAAACGATGACGCTGACTTAGCGCTGCTGGCGGTATGGCGCG 3752

QY 1621 ATGGGTGATGAGCGGTATGTA 1646

Db 3753 ATGGGTGATGAGCGGTATGTA 3778

RESULT 12
US-10-157-339-26
; Sequence 26, Application US/10157339
; Publication No. US20040005700A1
; GENERAL INFORMATION:
; APPLICANT: Surber, Mark W.
; APPLICANT: Matthew Giacalone
; TITLE OF INVENTION: POROPLASTS
; FILE REFERENCE: MPEX.008DV3
; CURRENT APPLICATION NUMBER: US/10/157,339
; CURRENT FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,951
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3840
; TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Fusion protein
US-10-157-339-26

Query Match 41.6%; Score 684.4; DB 12; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

QY 1 ATGGCTTCTAAAGAAATCCCTTTTGATGCTAAAGCCCGTGAAGAACTTTTCAGAGGTGA 60
Db 2133 ATGGCAGCTAAAGACGCTAAATTCGGTTAAGCAGCTCGTGTGAATATGCTGCGCGCGTA 2192

QY 61 GATAAATCTTGAATCTGTTAAAGTAAACACACTTGGACCTTAAAGCCGTAATCTGCTTAT 120
Db 2193 AAGCTACTGGCAGATGACGTGAAGTATACCTCGGTCCAAAGCCGTAACGTAAGTCTG 2252

QY 121 GAAAGCTTTTGGTCCCGAGTTATACAAAGATGGTGTATCTGTGCAAAAGAAAT 180
Db 2253 GATAAATCTTCCGTGACCGACCATACCAAGATGGTGTTCGGTGTCTCGTGAATC 2312

QY 181 GAATTTGAAGATAAGTTTGAATATGGCGCTCAAAATGGTTAAAGAGTAGTCCCAAA 240
Db 2313 GAATCGAAGACAAAGTTCGAAATATGGGTGCGCATGAAACCGTGAAGATGCTCTAA 2372

QY 241 ACTAGCGATATCTGCTGATGAACTACACAGACAGTCTTGCACAGCTATTTAT 300
Db 2373 GCAACGACGCTGCAGCGACCGTACACACATGCAACCGTACTGGCTCAGGCTATCATC 2432

QY 301 CGTGAAGGTGTAATACTTTAGCAGCTGGTGTGTAATCTTATGCCATTAACCGTGGCATA 360
Db 2433 ACTGAAGTCTGAAGCTGTGTGCGGCATGAAACCGTGAAGTGGTATC 2492

QY 361 GATAAAGCTGTGTTGCTGTTACTAAAGAACTAAGCGACATTAACAAGCCTACTCGTGAC 420
Db 2493 GACAAAGCGGTACCGCTGACGTGAAGAACTGAAAGCGCTGTCCGTACCATGCTGAC 2552

QY 421 CAATAAGAAATAGCTCAAGTTGAACCAATTTCTGCAACTCTGATACAACTAGTAAAT 480
Db 2553 TCTAAAGCGATTCTCAGGTGTGATCATCTCCGCTAACTCCGACGAAACCGTAGTAAA 2612

QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGAAAGAGGAGTGTATACAGTTGAGGAAGCT 540
Db 2613 CTGATCGCTGAAGCGATGGAACAAAGTCGGTAAAGAGCGGTATACCGTTGAAGACGGT 2672

QY 541 AAAGGCTTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
Db 2673 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTTCGACCGTGGCTACCTG 2732

QY 601 TCTCCATCTTTGTAATCTCTGAAATAATGGTTGTGAACCTTGATAACCTTATATC 660
Db 2733 TCTCCTTACTTATCAACAAGCCGGAACCTGGCGCAGTAGAAGTGGAAAGCCCGTTTATC 2792

QY 661 CTTTGTAAATCAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTTAGAACAGTT 720
Db 2793 CTGCTGCTGACAGAAAATCTCCACATCGCGAATATGTCGGTCTCTGGAAGCTGTT 2852

QY 721 GCTAAAGTAAACCGTCCACTCTTTATTTGCTGAAGACGCTAGAAGGTGAAGCACTTGCA 780
Db 2853 GCCAAAGCAGCAACCGCTGCTGATCATCGTGAAGATGTAGAAGCGGAGCGCTGGCA 2912

QY 781 ACATTGTAGTCAATAAGCTCCGTGGAGCACTCCAAAGTTGAGCCGTAAAGCTTCCTGGT 840
Db 2913 ACTCTGTTGTTTAAACACCATGCGTGGCATCGTGAAGTTCGTCGGGTAAAGCACCGGGC 2972

QY 841 TTTGGTCAACCGCTGAAGCTATCTTTGAAGATATTCCTTACTCTTACTGGAGGAGAACGA 900
Db 2973 TTGGCGCATCGTCTGAAGCTATCTGCGAGATATCGAACCCCTGACTGCGGTACCGTG 3032

QY 901 ATATTTGAAGATCGTGGTATAAAGCTTTGAAATGTAAAGCTTGTCTTCTTTAGGAACAGCT 960
Db 3033 ATCTCTGAAGAGATCGGTATGAGACTCGAAGAAAGCAACCCCTGGAAGACCTGGGTCAAGCT 3092

QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCTGTGATGGTCTGCTGAAATCAGAA 1020
Db 3093 AAACGTGTTGTGATCAACAAAGACACCACTATCATCATGCTGGTGGTGAAGACT 3152

QY 1021 GATATTAAAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAAACAGCTCAGATTAT 1080
Db 3153 GCAATCCAGGCGCGTGTGCTCAGATCCGTGAGCAGATTGAAGAACCACTTCTGACTAC 3212

QY 1081 GATCGTGAATAAATCTCAAGACGCTTTCGAAAACCTTGTGGTGGAGTAGCTGTATCCAT 1140
Db 3213 GACCGTGAATAAATCTCAGAACCGTAGCAAACTGGCAGCGCGGTTCAGATTATCAAA 3272

QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAAGAAAGGATCGCTAGAAAGATGCTCTA 1200
Db 3273 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAAAAGACGCGTTGAAGATGCCCTG 3332

QY 1201 AATGCAACAGAGCTCGGTTTGAAGAGGTATTGTCCTCGGTGGTGGTACTCTGTTTGTG 1260
Db 3333 CAGCGACCCGCTGCTCGGTGAGAAAGCGTGGTGTGCTGGTGGTGGTGGTGGTGGTGGT 3392

QY 1261 CGTCCATTAAAGTCTTGTATGATATTAAACCTGCTGATGATGATGAATCTGCTGACTT 1320
Db 3393 CGGTAGCGCTTAAACTGGCTGACCTGCGTGCAGAACGAGACGAGAACGCTGGGTATC 3452

QY 1321 AATATCATCCGCTGCTTCTTGAAGAGCCTTTACGTCAAATTTGCTCAAATGCTGGCTAT 1380
Db 3453 AAGTTGCACTGGTGCATGAAAGCTCCGCTCGCTCAGATCGTATTGAATCGCGGAA 3512

QY 1381 GAAGGTTCTATTGTTGTAGAAAAGTTGCTGAACCAAAAGATGGTTTGGATTTAATGCT 1440
Db 3513 GAACCGCTGTTGTTGCTAAACCCGTTAAAGGCGGACGCGCAACTACGCTTACAACGCA 3572

QY 1441 GCATGAGGAAATATGAAGACCTTATTAAGCTGTGCTGATTCATTCCTTAAAGTATACA 1500
Db 3573 GCAACCGAAGATACCGCAACATGTCAGATCGGATCTCTGGATCCACCAAGTAACT 3632

QY 1501 CGTATTGCATTACAAATGCAATGCAATCAGTACCTCTTACTTAACTACAGAAATGCGCT 1560
Db 3633 CGTTCTGCTGTCAGTACGAGCTTCTGCTGGCTGGCTGATGATCACCACCGATGATG 3692

QY 1561 ATTGCTGAAAACCCAGAACCTTAAAGATATGCTTATGCTGCGGTGGTGGTGGTGGT 1620
Db 3693 GTTACCGACCTGCGGAAAACGATGCACTTAGCGGCTCTGCGGTATGGCGGC 3752

QY 1621 ATGGGTGGTATGACCGTATGACTA 1646
Db 3753 ATGGGTGGCATGGCGGCATGATGA 3778

RESULT 13

US-10-157-305A-26
; Sequence 26, Application US/10157305A
; Publication No. US2003016609A1
; GENERAL INFORMATION:
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Surber, Mark W.
; APPLICANT: Neil Berkley
; APPLICANT: Anca M. Segall
; APPLICANT: Robert Klepper
; TITLE OF INVENTION: MINICELL COMPRISING MEMBRANE PROTEINS
; FILE REFERENCE: MPEX.008DV1
; CURRENT APPLICATION NUMBER: US/10/157,305A
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,951
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3840

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Gene encoding a fusion protein
US-10-157-391A-26

Query Match          41.6%; Score 684.4; DB 13; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conserved 0; Mismatches 601; Indels 0; Gaps 0;

QY 1 ATGGCTTCTAAGAAATCTTTTGTATGCTAAAGCCGCGGAAATCTTTCAGAGTGTA 60
Db 2133 ATGGCAGCTAAAGAGCTAAATTCGGTAACGACGCTCGTGTGAATATGTCGGCGGCTA 2192

QY 61 GATAACTTGAATGCTGTAAAGTAACACTTGAACCTTAAAGCCGCGTAATCTGTTAT 120
Db 2193 AACGTACTGGCAGATGCGAGTGAAGTTACCTCGGTCCAAAGGCCGTAACTAGTTCTG 2252

QY 121 GAAAAGTCTTTTGGTCCCGAGTTATTAACAAAGATGGTGTATCTGTGTCAAAAGAAAT 180
Db 2253 GATAAATCTTTCGGTGACCGACCATCACCAAGATGGTGTTCGGTTGCTGCGTGAATC 2312

QY 181 GAACCTGAAGATAAGTTTGAATAATATGGCGCTCAATGGTTAAAGAGTAGCTCCCAA 240
Db 2313 GAACTGGAAGACAAGTTTGAATAATATGGTGGCGAGATGGTGAAGAAGTTGCTCTAAA 2372

QY 241 ACTAGCGATATGCTGGTGAATGAACTACAACAGCAAGTCTTGCACAAAGCTATTAT 300
Db 2373 GAAACGAGCTGCGAGGCGACGTACCACTGCAACCGTACTGGCTCAGGCTATCATC 2432

QY 301 CTGAAGGTGTAAAACCTTGTAGCAGTGGTCTGTAATCTTATGGCCATTAAGCGTCATA 360
Db 2433 ACTGAAGGTCTGAAGCTGTGCTGGGCGATGACCCGATGACCTGAAACGTGGTATC 2492

QY 361 GATAAAGCTGTGTTGCTGTTACTAAAGAACTAAGCAATTAAGAGCTTACTCGTGAC 420
Db 2493 GACAAAGCGTTACCGCTGCACTGAAAGAACTGAAAGCGCTGTCGCTACCATGCTGAC 2552

QY 421 CAAGAAGAAATAGCTCAAGTTCGAACCAATCTGCAAACTCTGATACACATAGTAT 480
Db 2553 TCTAAGCGATGCTCAGTGGTACCATCTCCGCTAACTCCGCGAAACCGTAGGTAAA 2612

QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGAAAAGAGGTGTTATCACAGTTTGAAGAGCT 540
Db 2613 CTGATCGCTGAAGCGATGACAAAGTCTGTAAGAGCGGTATCACCGTTGAAGAGCT 2672

QY 541 AAAGTCTTGAACCTACATAGATGTTGTTGAAGAAATGAATTTGAACGCTGCTACCTC 600
Db 2673 ACCGCTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCACTGCAACCGTGGCTAC 2732

QY 601 TCTCCATCTTTGTAACTAATCTCAGAAAAATGTTTGTGAACTTGAATACCCCTTATATC 660
Db 2733 TCTCCTTACTTATCAACAGCCGGAACCTGCGCAGTAGTAAGTCTGAAAGCCCGTTCATC 2792

QY 661 CTTTGTAAAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTT 720
Db 2793 CTGCTGGCTGACAAGAAAATCTCAACATCCCGAAATGCTCCGCGTCTTGAAGCTGTT 2852

QY 721 GCTAAGCTAAACCGTCCACTCTTATATATGCTGAAGACGTAAGAGGTGAAGCACTTGA 780
Db 2853 GCCAAGCAGGCAAAACCGCTGCTGATCATCGTGAAGATGTAGAAGGCGAAGCGCTGGCA 2912

QY 781 ACACCTGTAGTCAATAAGCTCCGTGGACGACCTCAAGTTGTAGCCGTAAAGCTCCTG 840
Db 2913 ACTCTGTTTGTAAACACATGCTGGCATCTGTAAGTCTGCGTTTAAAGCACCGGCGC 2972

QY 841 TTTGTGAACCGCGTAAAGCTATGCTTGAAGATATTGCTTATCTTACTTGGAGGAGAAGCA 900
Db 2973 TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCTGACTGGCGGTACCGTG 3032

QY 901 ATATTGAGATCGGTGATATAAGCTTGAATAATGTAAGCTTGTCTTTTAGGAACAGCT 960
Db 3033 ATCTCTGAAGATCGGTATGAGCTGGAAGAAAGCAACCTTGAAGACCTGGGTGAGCT 3092
```

RESULT 14

```
US-10-157-391-26
; Sequence 26, Application US/10157391
; Publication No. US20030166279A1
; GENERAL INFORMATION:
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Neil Berkley
; TITLE OF INVENTION: MINICELL-BASED TRANSFECTION
; FILE REFERENCE: MPEX.008DV14
; CURRENT APPLICATION NUMBER: US/10/157,391
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,951
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3840
; TYPE: DNA
; ORGANISM: Artificial Sequence
```

FEATURE:
; OTHER INFORMATION: Fusion protein
US-10-157-391-26

Query Match 41.6%; Score 684.4; DB 13; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

```
QY 1 ATGGCTTTAAAGAAATCCCTTTTGTGATGCTAAAGCCCGTGAAGAAATCTTACAGAGTGTA 60
Db 2133 ATGGCAGCTAAAGACGTAAATTCGGTTACGACGCTCGTGTGAAAATGCTGCGCGCGTA 2192
QY 61 GATAAATCTTCAAAATGCTGTTAAAGTAACAATCTTGGACCTAAAGCCGCTAATGCTGTATT 120
Db 2193 AAGCTACTGGCAGATGAGTGAAGTTACCTCGGTCCAAAGCCGCTAAGTAGTTCG 2252
QY 121 GAAAGTCTTTTGGTTCCCGAGTTATTACAAAAGATGGTGATCTGTGTCAAAAGAAAT 180
Db 2253 GATAAATCTTTTCGGTGCACCGACCATCACCAAGATGGTGTTCCTCGTAAATC 2312
QY 181 GAATCTTGAAGATGAAGTTGAAAATATGGGCGCTCAAAATGGTTAAAGAGTAGCTCCAAA 240
Db 2313 GAATCTGGAACACAAAGTTTCAAAAATATGGGTGCGCAGATGGTGAAGAAAGTTGCTCTAAA 2372
QY 241 ACTAGCGATATTCTGTGTGATGAACTACAAACAGCAACAGCTCTTGCACAAAGCTATTAT 300
Db 2373 GCAACGACGCTGAGCGACGGTACCAACACTGCAACCGTACTTGGCTCAGGCTATCATC 2432
QY 301 CGTGAAGGTGTAAACTTTGAGCAGCTGGTCTGAATCCTATGGCCATTAACAGCTGGCATA 360
Db 2433 ACTGAAGTCTGAAGCTGTGTGCGGGCATGAACCCGATGGACCTGAACAGTGGTATC 2492
QY 361 GATAAGCTGTGTGTGTTACTTAAAGAACTAAGCGACATTAAGAGCCTACTCTGTGAC 420
Db 2493 GACAAAGCGGTATACCGTGCAGTTGAAGAACTGAAAGCGGTGTCGCTACCATGCTCTGAC 2552
QY 421 CAAGAAATAGTCAAGTTGGAACCATTTCTGCAAACTCTGATACAAACAAATAGGTAAAT 480
Db 2553 TCTAAAGCGATGTCTAGGTGTGACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 2612
QY 481 ATCATAGCTCAAGCTATGGCTTAAAGTTGGAAGAGGTGTATCAAGTTGAGGAAGCT 540
Db 2613 CTGATCGCTGAGCGATGGAAGTGTGCTGCGGCATGAACCCGATGGACCTTATCACCCTTGAGACGGT 2672
QY 541 AAGGCTTGAAGTCACTATGATGCTGGTGAAGGAATGAAGTTGACCGTGGCTACTCTC 600
Db 2673 ACCGGTCTGACGACGAACTGGAAGTGTGAGGTATGCAAGTATGCAAGCTGCGCTGACCTG 2732
QY 601 TCTCCATCTTTGTAACCTAATCTCGAAGAAATGTTTGTGAACCTTGATAACCTTATATC 660
Db 2733 TCTCCTTACTTTCATCAACAGCCGGAACCTGGCGCAGTAGAAGTGAAGCCGCTTCAATC 2792
QY 661 CTTTGTAAATGAGAAAAAGATTACTAGCATGAAAGACATGCTTACCAATCTTTAGAACAAAGTT 720
Db 2793 CTGCTGCTGACAGAAATCTCAACATCCGCGAAATGCTGCGGTCTTGTGAAGCTGTT 2852
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATGCTGAAGACGTAGAGGTGAAGCACTTGCA 780
Db 2853 GCCAAAGCAGCAAAACCGCTGCTGATCATCTGAAGATGTAGAAGCGCAAGCGCTGGCA 2912
QY 781 ACATCTTGTGTCATAGCTCCGTTGAGGAGCACTCCCAAGTTGAGCGGTAAAGCTCTTGGT 840
Db 2913 ACTCTGGTTGTAAACCATCCGTTGGCATCTGTGAAGTTCGCTGGTTTAAAGCAACCGGGC 2972
QY 841 TTTTGGTGAACCCGCTAAAGCTATGCTTGAAGATATTGCTATCTTACTTCTAGGAGAGAACA 900
Db 2973 TTTGGCGATCTGCTGAAGCTATGCTGCAGATATGCAACCCCTGACTGCGGTACCGTG 3032
QY 901 ATATTTGAAGATCGGTGATTAAGCTTGAAGATGTAAGCTTGTCTTCTTTAGAACAGCT 960
Db 3033 ATCTCTGAAGAGATCGGTATGAGCTGGAAGAAAGCAACCCCTGGAAGACCTGGTCAAGCT 3092
QY 961 AAACGTGTAGTTATTGACAAAGAAAATACTACTACTCTGTGATGGTGTGGAATAATCAGAA 1020
```

```
Db 3093 AAACGTGTTGTGATCAACAAAGACACCACTATCATCGATCGCTGGGTGAAGAAGCT 3152
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCAGAAACCAAGCTCAGATTAT 1080
Db 3153 GCAATCCAGGGCGGTGTGCTCAGATCCGTCAGAGATTGAAGAAGCAACTTCTGACTAC 3212
QY 1081 GATCGTGAAGAACTTCAAGAAACGCTTTGCAAAACTTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db 3213 GACCGTGAAGAACTGCAAGAACGCTAGCAAACTGGCAGCGCGCTTGCAGTTATCAAA 3272
QY 1141 GTTGGAGCTGCTACTGAAACTGAAATGAAGAAAGAAAGGATCGTGPAGAGATGCTCTA 1200
Db 3273 GTGGGTGCTGCTACCGAAAGTTGAAATGAAGAAAGAAAGACGCGTTGAAGATGCCCTG 3332
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCCTGGTGGTGTACTGTTTGTGTC 1260
Db 3333 CACGCAACCGCTGCTCGGTAGAAAGGCGTGGTGTGCTGGTGGTGTGTTGCGCTGATC 3392
QY 1261 CGCTCCATTAAAGCTCCTTGTGATGATATTAAACCTGCTGATGATGATGAACCTTGTGGACTT 1320
Db 3393 CGGTAGCTGTAAACTGGCTGACCTGCGTGGTCAAGAACCAAGACGTTGGGTATC 3452
QY 1321 AATATCATCCGTCGTTCTTCTTGAAGAGCCCTTTACGTCAAATTCGTGCAAAATTCGTGCTAT 1380
Db 3453 AAGTTGCACTGCTGCTGCAATGGAAGCTCCGCTCGCTCAGATCGTATTGAATCGCGCGAA 3512
QY 1381 GAAGGTTCTATTGTTAGAAAAGTTCTGTGAACCAAGATGCTGTTTGAATTAATGCT 1440
Db 3513 GAACCGTCTGTTGTTGCTAACCCGTTAAAGCGCGGCAACGCTTACAAAGCA 3572
QY 1441 GCATCAGGAAATGAAGACCTTATTAAAGCTGGTGTCTATTGATCCTAAAAAGTTTACA 1500
Db 3573 GCAACCGAAGATACCGCAACATGATCGACATCGGTATCTCTGATCAACCAAGTAACT 3632
QY 1501 CGTATTGCATTAACAAATGCAAGATGCAAGTACGTAGCTCTCTTACTTCTAACTACAGAAATGCGCT 1560
Db 3633 CGTTCTGCTCTGAGTACGAGCTTCTGTGGTGGCTGATGATCACCAAGATGATG 3692
QY 1561 ATTGCTGAAAACCAAGAACCTTAAAGATATGCTATGCTGCGGTGTTGGTGGT 1620
Db 3693 GTTACCGACCTCGCAAAACGATGCACTGACCTAGGCTGCTGCGGTATGCGCGGC 3752
QY 1621 ATGGGTGATGACCGTATGTA 1646
Db 3753 ATGGGTGCAATGGCGCATGATGA 3778
RESULT 15
US-10-157-096-26
; Sequence 26, Application US/10157096
; Publication No. US20030190601A1
; GENERAL INFORMATION:
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Neil Berkley
; APPLICANT: Surber, Mark W.
; TITLE OF INVENTION: TARGET DISPLAY ON MINICELLS
; FILE REFERENCE: MPEX.008DV12
; CURRENT APPLICATION NUMBER: US/10/157,096
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,51
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3840
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

OTHER INFORMATION: Fusion protein
US-10-157-096-26

Query Match 41.6%; Score 684.4; DB 13; Length 3840;

Best local Similarity 63.5%; Pred. No. 6.6e-130;

Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

```
QY 1 ATGGCTTCTAAAGAAATCCCTTTTGATGCTAAAGCCCGTGAAGAACTTTTACAGAGGTGA 60
DB 2133 ATGGCAGCTTAAGACGTAAATTCGTAAACGACGCTCGTGTGNAATGCTGCGGCGTA 2192
QY 61 GATAAAGTCTGAAATGCTGTTAAAGTAAACACTTGAACCTTAAAGCCGTAATGCTGTTAT 120
DB 2193 AAGCTACTGCGCAGATGCAAGTAAAGTTACCTCGGTCCAAAGGCGGTAAACGTAGTTCTG 2252
QY 121 GAAAGTCTTTTGGTTCCCAAGTATTACAAAGATGTTGATCTGTTTCCAAAGAAAT 180
DB 2253 GATAAATCTTTTCGGTGCACCGACCATCACCAAGATGTTGTTTCGGTTCTCGTGAATC 2312
QY 181 GAACTTGAAGTAAGTTTGAAATATAGGCGGCTCAAAATGGTTAAAGAGTAGTCCCAAA 240
DB 2313 GAACTGGAAGACAAGTTTCAAAATATAGGTCGACAGATGTTGAAGAGTTGCTCTAAA 2372
QY 241 ACTAGCGATATTGCTGGTGAAGTGAACCTAACAGCAACAGTCTCTTGCAACAAGCTATTAT 300
DB 2373 GCAAAACGACGCTGCAAGGCGACGCTACCACTGCAACCGTACTGGCTCAGGCTATCATC 2432
QY 301 CGTAGAGTGTAAACTTGTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 2433 ACTGAAGTCTGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2492
QY 361 GATAAGCTGTTGTTGCTTAAAGAACTAAGCGACATTAACAAAGCTTACTGCTGAC 420
DB 2493 GACAAAGCGTTTACCGTGCAGTTGAAGAACTGAAAGCGCTGTCGTTACCATGCTGAC 2552
QY 421 CAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAACTCTGATGATGATGATGATGATG 480
DB 2553 TCTAAGCGATGCTCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2612
QY 481 ATCATAGCTGAAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 2613 CTGATGCTGATGAGCGATGACAAAGTCGTTGAAGAGGCTGTTATCACCGTTGAAGCGGT 2672
QY 541 AAGGCTTTGAACTACATTAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 2673 ACCGCTCTGCAGGACGAAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2732
QY 601 TCTCATATCTTTGTAATCTCTGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 2733 TCTCCTTACTTCAACAAGCGCGAAACTGCGCGAGTAGAAGTGAAGCGCGGTTCATC 2792
QY 661 CTTTGTATGAGAAAGATTAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 2793 CTGCTGGCTGACAGAAATCTCAACATCCGCGAAATGCTGCGGTTCTGGAAGCTGTT 2852
QY 721 GCTAAGTAAACCGTCCACTCTTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 2853 GCCAAGCAGCAACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2912
QY 781 ACATTTAGTCAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 2913 ACTCTGTTGTTAACCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2972
QY 841 TTTGTGTAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 2973 TTCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3032
QY 901 ATATTGAGATCGTGGTAAAGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 960
DB 3033 ATCTCTGAAGAGATCGTGGTGAAGCTGGAAGAAAGCAACCTGGAAGACCTGGTCAAGCT 3092
QY 961 AAACGTGATGTTATTGACAAAGAAATACTACTATCTGTTGATGCTGCTGCTGCTGCTGCT 1020
```

```
DB 3093 AAACGTGTTGCTGATCAACAAAGACACCAACCATATCATCGATGGCTGGTGAAGAGCT 3152
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATTAT 1080
DB 3153 GCAATCCAGGCGCGTGTCTCAGATCCCTCAGATCCCTCAGATTTGAAGAAACAACTTCTGACTAC 3212
QY 1081 GATCGTGAAGAACTTCAAGAAAGCTTTCGAAACCTTGTGGTGGAGTAGCTGTTATCCAT 1140
DB 3213 GACCGTGAAGAACTCGAGAAAGCTGAGGAAAGCTGCGAGGCGCGTTCAGTATATCAA 3272
QY 1141 GTTGAGCTGCTTACTGAAACTGAAATGAAAGAGAGAGAGATGCTGTAGAAGATGCTCTA 1200
DB 3273 GTGGTGTCTGCTACCGAAGTTGAAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3332
QY 1201 AATGAAACAGAGCTGCGGTTGAAGAGATTTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 3333 CAGCGACCCGCTGCTGCGGTGAAGAGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3392
QY 1261 CGCTCCATTAAAGTCTTGTGATGATATTAAACCTGCTGCTGATGATGATGATGATGATGAT 1320
DB 3393 CGCGTAGCTCTAAACTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3452
QY 1321 AATATCATCCGCTGCTTCTTGAAGAGCTTTTACGTCAAAATTTGCTGCAAAATGCTGCTAT 1380
DB 3453 AAAGTTGCACTGCTGCTCAATGAAAGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3512
QY 1381 GAAGTTCTTATGTTGTAGAAAGATTTGCTGAAACCAAGAGATGCTTTCGATTTATGCT 1440
DB 3513 GAACCGTCTGTTGTTGCTAAACACCGTTAAAGCGGCGAGCGCAACTACCGTTTACAAACGA 3572
QY 1441 GCATCAGAGAGATATGAAGACCTTATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
DB 3573 GCAACCGAAGATACGCAACATGATCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3632
QY 1501 CGTATTGCAATTAACAAATGCAAGATGCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
DB 3633 CGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3692
QY 1561 ATTGCTGAAACACCAAGACCTTAAAGAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
DB 3693 GTTACCGACCTCCGAAACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
QY 1621 ATGGGTGCTATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
DB 3753 ATGGGTGCTATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
```

Search completed: January 29, 2004, 16:43:39
Job time : 485 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 13:07:12 ; Search time 89 seconds
(without alignments)
8168.068 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 1647

Sequence: 1 atggctcttaagaatactct.....gtatgaacggtatgtagtag 1647

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A COMB.seq.*

2: /cgn2_6/prodata/2/ina/5B COMB.seq.*

3: /cgn2_6/prodata/2/ina/6A COMB.seq.*

4: /cgn2_6/prodata/2/ina/6B COMB.seq.*

5: /cgn2_6/prodata/2/ina/PCRUS_COMB.seq.*

6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	748.8	45.5	1653	US-09-328-352-1382	Sequence 1382, Ap
2	725.8	44.1	1830121	US-09-557-884-1	Sequence 1, Appli
3	725.8	44.1	1830121	US-09-643-990A-1	Sequence 1, Appli
C 4	690.2	41.9	1230025	US-09-198-452A-1	Sequence 1, Appli
5	689.8	41.9	1926	US-09-613-303-50	Sequence 50, Appli
6	687.6	41.7	5365	US-08-961-527-77	Sequence 77, Appli
7	682.8	41.5	1647	US-09-472-971-5	Sequence 5, Appli
8	682.8	41.5	4524	US-09-472-971-7	Sequence 7, Appli
9	642.4	39.0	1838	US-08-470-260-7	Sequence 7, Appli
10	642.4	39.0	1838	US-08-471-491-7	Sequence 7, Appli
11	642.4	39.0	1838	US-08-466-662-7	Sequence 7, Appli
12	642.4	39.0	1838	US-08-256-847C-6	Sequence 6, Appli
13	640	38.9	1623	US-09-134-001C-1868	Sequence 1868, Ap
C 14	626.4	38.0	2416	US-09-221-017B-895	Sequence 895, App
15	610.4	37.1	2284	US-08-467-822-28	Sequence 28, Appl
16	610.4	37.1	2284	US-08-432-697-28	Sequence 28, Appl
17	610.4	37.1	2284	US-08-466-248-28	Sequence 28, Appl
18	553.2	33.6	1656	US-09-252-991A-7638	Sequence 7638, Ap
C 19	553.2	33.6	1722	US-09-252-991A-7917	Sequence 7917, Ap
20	480.6	29.2	1050	US-09-107-532A-3030	Sequence 3030, Ap
21	441.2	26.8	1777	US-08-828-199A-1	Sequence 1, Appli
22	440	26.7	580073	US-08-545-528B-1	Sequence 1, Appli
C 23	439.8	26.7	2051	US-09-207-388-8	Sequence 8, Appli
24	439.4	26.7	1647	US-09-207-388-9	Sequence 9, Appli
25	439.4	26.7	1707	US-09-207-388-10	Sequence 10, Appl
26	427.4	26.0	1653	US-09-207-388-6	Sequence 6, Appli
27	427.4	26.0	1713	US-09-207-388-7	Sequence 7, Appli

28	425.8	25.9	1944	4	US-09-613-303-52	Sequence 52, Appli
29	406.2	24.7	1623	4	US-09-613-303-3	Sequence 3, Appli
30	406.2	24.7	1920	4	US-09-613-303-16	Sequence 16, Appli
31	406.2	24.7	1947	4	US-09-613-303-28	Sequence 28, Appli
32	406.2	24.7	2847	4	US-09-613-303-20	Sequence 20, Appli
33	406.2	24.7	4380	4	US-08-955-565A-3	Sequence 3, Appli
34	406.2	24.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
35	406.2	24.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
36	396.6	24.1	1620	2	US-08-461-775-10	Sequence 10, Appli
37	396.6	24.1	1620	3	US-09-031-606-10	Sequence 10, Appli
38	390.8	23.7	1569	2	US-08-997-080-113	Sequence 113, App
39	390.8	23.7	1569	2	US-08-997-362-113	Sequence 113, App
40	390.8	23.7	1569	3	US-09-095-855-113	Sequence 113, App
41	390.8	23.7	1569	4	US-09-324-542-113	Sequence 113, App
42	390.8	23.7	1569	4	US-09-205-426-113	Sequence 113, App
43	390.8	23.7	1626	2	US-08-997-080-159	Sequence 159, App
44	390.8	23.7	1626	2	US-08-997-362-159	Sequence 159, App
45	390.8	23.7	1626	3	US-09-095-855-159	Sequence 159, App

ALIGNMENTS

RESULT 1

US-09-328-352-1382

; Sequence 1382, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 1382

; LENGTH: 1653

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-1382

Query Match 45.5%; Score 748.8; DB 4; Length 1653;

Best Local Similarity 66.6%; Pred. No. 2.2e-187;

Matches 1087; Conservative 0; Mismatches 542; Indels 3; Gaps 1;

QY	1	ATGGCTTCTAAGAAATCCCTTTTGTGATGCTAAAGCCGCTGAAACCTTTCCACAGGTGTA	60
DB	19	ATGTCAGCTAAAGACGCTAAATTTGGTGTTCAGCTCGCTCAAAAATGATTGCGAGGGTA	78
QY	61	GATAAACTTGCAATGCTGTTAAAGTAACACTTGGACCTAAAGCCGCTAATGTCGTATT	120
DB	79	ACGTACTTGCAGATCGGTTAAAGTACCTTTAGGTCCTAAAGCCGCTAATGTTGTATC	138
QY	121	GAAGAGCTTTTGGTTCCCGAGTTATTACAAAAGATGGTGTATCTGTGCAAGAAAT	180
DB	139	GACCGCTCTTTGCGCGCGCGCACATCCTAAAGACGGTGTACTGTGCAAGAAAT	198
QY	181	GAACCTGAAGATTAAGTTGAAATATGGCGCTCAATGTTAAGAGTAGTCCCAAA	240
DB	199	TCATTAAGACAAAGTTTGAACCATGGTGTCTCAACTGTTGCTGAAGTTTCAAGCAA	258
QY	241	ACTAGCGATATTGCTGCTGATGGAACCTACAGCAACAGTCTCTTGCACAAGCTATTAT	300
DB	259	ACTAAGCATCCGAGTGACGGTACCAACTGCAACTGCTACTTCTCAAGCAATTTTA	318
QY	301	CGTGAAGGTGTAACCTTTAGCAGCTGTGCTGTAATCTTATGGCCATTAAACCTGGCATA	360
DB	319	ATGAAGGTATCAATCACTGCTGCTGATGAACCCCAATGATTTAAACCGGTATC	378
QY	361	GATAAGCTGTTGTTGCTGTTACTTAAAGAACTAAGCGACATTACAAAGCCTACTCGTAC	420
DB	379	GACATTGCAGTAAACACTGTAGTTGAAATATCCGTTCTTCTTAAACACCGCTGATGAT	438

QY 421 CAAAGAGAAATAGCTCAAGTTGGAAACCAATTTCTGCAAACTCTGTATACAACTAGGTAAT 480
Db 439 TTCAAGCAATTTGAACAAGTAGTTCAATCTCTGCTAACTCTGATCTACTGTTGGTAA 498
QY 481 ATCATAGCTGAAGCTATGCTAAAGTTGGAAGAGAGGTGTTATCACAGTTTGAGGAAGCT 540
Db 499 CTTATTTCTCAAGCAATGGAAAGAGTAGTTAAAGAGCGGTAATCACTGTAGAGAGGC 558
QY 541 AAAGGTCTTGAACATACATAGATGTTGTTGAAGGAATGAAGTTTGACGGTGGCTACCTC 600
Db 559 TCTGCTTCGAAGACGCAATAGACCTTGTGAAGGTATGCAAGTTTGAAGCGTGGTTATATC 618
QY 601 TCTCATATCTTTGTAACCTCTCAGAAATGTTTGTGAACCTTGAATACCTTATATC 660
Db 619 TCTCGTACTTTGCAACAACAAGATATCTTAACCTGCTGAACCTTGAATACCTGTTATC 678
QY 661 CTTTGTATGAGAAAGATTAAGCTATGATGAAGAGATGCTACCAATCTTGAACAAGTT 720
Db 679 CTTCTTGTGATAAAAAATCAGCAATCTGTAATGATTTCTGTTTGAAGCAGTT 738
QY 721 GCTAAAGTAAACCTGCTCCTCTTATTTATGCTGAAGAGTAGAGGTGAAGCACTTGA 780
Db 739 GCTAAACTGCTAAACCTCTTATCATCTGCTGAAGATTTGAAGGTGAAGCGCTTGA 798
QY 781 ACACCTGTAGTCAATAAGCTCGTGGAGCACTCCAAAGTTGTAGCCGTAAAGCTCCTGCT 840
Db 799 ACTCTGTGTGAACAACAATGCGCGGTATCATCAAGTATGCTGTTTGAAGCTCCTGCT 858
QY 841 TTTGTGAACGCTGTAAGCTATGCTTGAAGATATTTGCTATCTTACTTACTGGAGGAGA 900
Db 859 TTTGCTGACCGCTGTAAGCAATGCTTCAAGCAATTTGCTGCTGCTGCTGCTGCTGCT 918
QY 901 ATATTTGAAGATCGTGGTATTAAGCTTGAAGTGAAGCTTCTCTTTTAGAACAGCT 960
Db 919 ATTTCTGAAGAGTTGGTATGCTTTAGAACAGCACTCTTCAAGATTTAGGTACAGCG 978
QY 961 AAACGTGTAGTTATTGACAAAGAAATATCTATCTGTTGATGCTGCTGGAATAACAGAA 1020
Db 979 CACAAATTTACTTTCTTAAGAAACACAGTTATTTGAGCGTGTGCTGCTGCTGCT 1038
QY 1021 GATATTAAGCTGAGTTTAAACAAATTTGCTGACAAATTTGAAGAAACAGCTCAGATTAT 1080
Db 1039 GCTATCGCTGAGCGTGTCAACAAATTCGTGCTCAAAATCGAAGAAATCTACTTCAGAA 1098
QY 1081 GATCTGGAAGAACTTCAAGAAAGCTTGTCAAAAGCTTGTGTTGGAGTAGCTGTTATCAT 1140
Db 1099 GACCGTGAAGAAATTAAGAAAGCTGTTGCTAAGTTAGCAGCGGTGTTGCTGTAATTA 1158
QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAGAGAGAGGATCGTGTAGAAGATGCTCTA 1200
Db 1159 ATCGTGCAGCACTGAACTGAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1218
QY 1201 AATGCAACAGAGCTGCGGTGGAAGAGATATGTCCTCGGTGGTGGTACTGCTTTGTC 1260
Db 1219 CACGCAACTCGTGCAGCAGTGAAGAGAGTGTGTTGCTGGTGGTGGTCTGCTCTGTT 1278
QY 1261 CGCTCCATTAAGCTCTTCAATGATATTAACCTGCTGATGATGATCACTTCTGCTGACT 1320
Db 1279 CCGCTGTAAATGCTCTTGAAGGTTTAAAGGCGCTTAAGAGAGTAAACAGCTGGTATC 1338
QY 1321 AATATCATCGCTGCTTCTTGAAGAGCGCTTTACGTCAAAATGCTGCAAAATGCTGCTAT 1380
Db 1339 ACATTTTACGCGGTGCGATCGAGCTCACTTCGTCAATCGTTGCGAATGCTGGTAT 1398
QY 1381 GAAGGTCTTATTTGTGAGAAAGTTCTGTGAACCAAGATGTTTGGATTTAATGCT 1440
Db 1399 GAGCCATCTGTAGTTATCAATGCTTGAAGATGTTGAGGGTAACTTCCGGCTACAAATGCT 1458
QY 1441 GCATCAGGAGATATCAAGCACTTATTAAGCTGCTGCTATTAAGCTTCAAAAGTTACA 1500
Db 1459 GCAACTGCGCAATATGCTGATATGTTAGAAATGGGTATCCTTGACCCAGCTAAAGTACT 1518
QY 1501 CGTATTTGCATTACAAATGCAGCATCAGTAGCTCCTTACTTCTTAAGTACAGAAATGCGCT 1560

Db 1519 CGTTCTGCACCTTGAGCACGCTGCTTCTGTTGCTGCTTAATGTTAACTACAGAAATGATG 1578
QY 1561 ATTGCTGAAACACAGAACCTTAAAGAAATATGCTATGCTGCGGTGGTATGCTGCTG 1620
Db 1579 ATTATGATCTTCTCTGAAGACAAACCACTGCTGCTCAGATATGCGGCTGCTGCTG 1635
QY 1621 ATGGTGTGATG 1632
Db 1636 ATGGCGGAATG 1647
RESULT 2
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506591
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 44.1%; Score 725.8; DB 4; Length 1830121;
Best Local Similarity 65.9%; Pred. No. 78-180;
Matches 1086; Conservative 0; Mismatches 557; Indels 6; Gaps 2;
QY 1 ATGGCTTCTAAAGAAATCTTTTGTATGCTAAAGCCCGTGAAGAACTTTTACGAGGTGA 60
Db 565350 ATGGCAGCAAGAGACGTAATTTGGTAATGACGACGCGTAATGCTTTAAGCGCTG 565409
QY 61 GATAAATCTGCAATCTGTTTAAAGTAACACTTGAACCTTAAAGCGCTAATGCTGCTAT 120
Db 565410 AATGATTAGCGGATCGATAAAGTAACCTTTGGCCCGAAAGGCTGCTGTAATTTTA 565469
QY 121 GAAAGTCTTTTGGTTCCCGCTATTTACAAAGATGTTGTTATCTGTTCAAAAGAAAT 180
Db 565470 GATAAATCATTTGGCGCACCACTATCAAAAGACGGTGTGTTCTGTTGCTGTAATC 565529


```

; LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature

Query Match 41.9%; Score 690.2; DB 4; Length 1230025;
Best Local Similarity 65.1%; Pred. No. 1.3e-170;
Matches 1035; Conservative 0; Mismatches 549; Indels 6; Gaps 1;

QY 1 ATGGCTCTTAAGAAATCCTTTTGATGCTAAAGCCGCGGAAAGAACTTTTCAGAGGTGA 60
DB 179472 ATGGCAGCGAAATAATTAATATATGAAGAGCGCAGANAAAAATACATAAGGGGTA 179413

QY 61 GATAAATCTGCAATCTCTTAAAGTAACACATTTGACCTTAAAGCCGCTAATGTCGTTATT 120
DB 179412 AAAATCTTTCGACAGCAGTAAAGTTACTCTAGTCTCTAAAGGAGTCACTAGTTATA 179353

QY 121 GAAAGTCTTTGGTTCCCGAGTTATTAACAAGATGGTGTATCTGTTCCAAAGAAAT 180
DB 179352 GATAAGAGCTTTGGCTCTCCCAAGTGAATGATGTTTACTCTAGCTAAAGAAATC 179293

QY 181 GAACCTGAAGTAAGTTTGAATAATGGGCGCTCAATGTTAAAGAGTAGTCCCAA 240
DB 179292 GAGCTCGAAGACAAACATGAAACATGGGCGCTCAGATGGTAAAGAGTCCGACAAA 179233

QY 241 ACTAGCGATTTCTGCTGATGGAATACAAACAGCAACAGTCTCTGCAAGCTATTAT 300
DB 179232 ACTGCTGACAAAGCAGCGACGAACTACAAAGCACTGTTCTGCAAGCAATCTAT 179173

QY 301 COTGAAGGTGTAACCTTTAGCAGTGTGTAATCTGATGCGCAATTAACGTTGCATA 360
DB 179172 AGCGAAGGTCTAAGAAATGTCATGCGCGTGGCAATCTCTATGGACCTAAAGAGGTATC 179113

QY 361 GATAAGCTGTTCTGCTGTTACTAAGAACTAAGCAGATTAACAAAGCTACTCTGAC 420
DB 179112 GACAAGCGCTAAAAGTTGTTGTTGATGAACCTCAAAAAATTAGTAAACCTGTACAACAT 179053

QY 421 CAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATAACAATAGGTAAT 480
DB 179052 CACAAGAAATCGCTCAAGTAGTACTATCTCAGCAAAATATGATTCGGAATCGAAT 178993

QY 481 ATCATAGCTAAGCTATGGCTAAAGTTGGAAGAGGAGGTGTTATCAAGTTGAGGAGCT 540
DB 178992 CTTATTGCAAGAGCTATGAAAAAAGTTGGTAAAAACGGATCCATTACTGTTGAAGAGCT 178933

QY 541 AAAGGTCTTGAATACATTAGATGTGGTTGAAGAAATGAAGTTTGAACGCTGCTACCTC 600
DB 178932 AAAGGCTTCGAAATCTGTTCTCGAGTTGTAGAGGATGAATCTTCAACGCTGATACCTC 178873

QY 601 TCTCCATCTTTGTAACTAATCTCAGAAAAATGGTTTGTGAACCTTGATAACCTTATATC 660
DB 178872 TCCAGCTACTTCTCCACAAATCCAGAACTCAAGAATGGTTTGAAGACGCTCTGATT 178813

QY 661 CTTTGTATGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
DB 178812 CTAATCTAGATAAAAAATCTCTGCAATTAAGACTTCTTCCAGTTTGAACAAGTA 178753

QY 721 GCTAAAGTAAACCGTCCACTCTCTTATTGCTGTAAGACGTAGAAGTGAAGCACTTGA 780
DB 178752 GCAGAACTGGAAGCGCTCTTTTATCATTCAGAGAANAATGAAGGAGAGCTTTAGCA 178693

QY 781 ACACCTGTAGTCAATAAGCTCGGTGAGCACTCCAAGTTGTAGCCGCTAAAGCTCTCTGT 840
DB 178692 ACTCTAGTAGTCAATAGACTCGGTGAGGATTCAGAGTCTGTGCAAGTGAAGCTCTCTGT 178633

QY 841 TTTGGTGAAGCGCTAAGCTATGCTTGAAGATATTGCTATCTTACTTCTTCTTCTTCTT 900
DB 178632 TTCGGTGAAGAGAAAGCTATGTTAGAGACATCGCTATCTTACTGCTGCGCCAACTA 178573

QY 901 ATATTGTAAGATCGTGGTATAAAGCTTGAAGATGTAAGCTTCTTCTTCTTCTTCTTCTT 960
DB 178572 GTTAGCGAAGAACTTGGCATGAACCTAGAGATACAACTCTAGCAATGTTAGGAAGCT 178513

QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCTGTTGATGGTCTGGAATAACAGAA 1020
DB 178512 AAGAAAGTTTATCGTAACATAAGAGATACCAATCTGCGAAGCTTAGGAAACAACT 178453

QY 1021 GATATTAAAGCTCGAGTTAAACAAATCTGTCACAAATTTGAAGAAACAGCTCAGATTAT 1080
DB 178452 GATATCCAGCTCGATGCGACAAATATTAAGAAACAAATCGAAGATAGCACTCAGATTAC 178393

QY 1081 GATCGTGAAGAACTTCAAGAACTCTTGAAGAACTTGTGGTGGAGTAGCTGTATCCAT 1140
DB 178392 GACAAGAAACCTCAAGAGCGTTTGTAGTAACTCTCGGTGGTGTGCGCGTAAATCCGC 178333

QY 1141 GTTGAGCTGTCTACTGAACTGAAATGAAAGAGAGAGAGATCGTGTAGAGATGCTCTA 1200
DB 178332 GTAGAGCTGTCTACCGAAATAGAGATGAAAGAGAAAGAGACAGATAGATGACAAA 178273

QY 1201 AATGCAACAGAGCTGCGGTTCGAAGAGTATTGTCCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 178272 CACGCAACCATTTGCAAGTGTGCAAGAGGAATCTCTCTGCTGCTGCTGCTGCTGCTG 178213

QY 1261 CGCTCCATTAAAGTCTCTGATGATATTAACCT-----GCTGATGATGATGAACCTGCT 1314
DB 178212 CGCTGATCCCTACACTAGAGCTTCTCTCTATGCTAGCAACGAGAGAGAGCTATT 178153

QY 1315 GGACTTAATATCATCCGTCGTTCTCTTGAAGAGCTTTTACGTCAAAATTCGCGAAATGCT 1374
DB 178152 GGTACTCGTATTATTCTAAAGCAATTAACAGCTCCATTAAGCAAAATTCGAAGTAAAGCA 178093

QY 1375 GCTATGAGGTCTTATGTTGTTAGAAAAGTTGTTGTAACAAAAGATGGTTTGGATT 1434
DB 178092 GGTAAAGAGGCGCTATCATTTGTGAGCAAGTTCTAGCAAGATCTGCAAAATGAAGCTAT 178033

QY 1435 AATGCTGCATCAGGAGATATGAAGCACTTATAAGCTGGTGTCTATGATCTCAAAAAA 1494
DB 178032 GATGCTTACGTGACGCTTATACAGATGATGTTGACGAGGAATTTTAGTCCAATAA 177973

QY 1495 GTTACAGCTATTGCAATTAACAAATGCAGATCAGTAGCTCTTACTTTCTAATAACAGAA 1554
DB 177972 GTGACTCGCTCAGCTAGAAAAGCGAGCTTCTATCGAGGATTACTCTCTCAACAGAA 177913

QY 1555 TCGGCTATTGCTGAAAACCCAGAACCTTAA 1584
DB 177912 GCCTTAATCGCTGATATCCAGAGAGAAA 177883
```

RESULT 5

US-09-613-303-50
; Sequence 50, Application US/09613303
; Patent No. 6495347
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/09/613,303
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
; NAME/KEY: CDS
; LOCATION: (1)...(1923)
US-09-613-303-50

Query Match 41.9%; Score 689.8; DB 4; Length 1926;

Best Local Similarity 65.8%; Pred. No. 7.7e-172;
Matches 1035; Conservative 0; Mismatches 532; Indels 6; Gaps 2;

QY	8	CTAAGAAATCCCTTTTGTATGCTAAAGCCGCTGAAATCTTTCACGAGGTGTAGATAAC	67
DB	5	CAAGAAATTAATTTTCATCAGATCCCGTTCAGCTATGTCCTGGTGCATATCC	64
QY	68	TTGCAATGCTGTTAAAGTAACTTGGACCTAAAGCCGCTAAATGCTGTTATGAAAGT	127
DB	65	TTGCAATGCTGTTAAAGTAACTTGGACCTAAAGCCGCTAAATGCTGTTATGAAAGT	124
QY	128	CTTTTGGTCCCGAGTTTACAAAGATGCTGATCTGTTGCAAAAGATTAAGTAACTG	187
DB	125	CATTGCTGCTCCCTTGAATTAACCAATGACGCTGTGACTATTGCAAAAGATTAAGTAA	184
QY	188	AAGATAGTTGAAATATGGCGCTCAATGTTAAAGAGTAGTCCCAAACTAGCG	247
DB	185	AAGACCAATTTTGAATATGGTGCCAAATGGTATCAGAGTAGTCTCAAAACCAATG	244
QY	248	ATATTGCTGCTGATGGAACCTACAAAGCAAGCTCTTGGACAGCTATTATCGTGAAG	307
DB	245	ATATCGCAGGTGATGGAACCTACAAAGCAAGCTCTTGGACCAAGCAATCGTCCGTGAAG	304
QY	308	GTGTAAACTTTAGCAGCTGGTGTATCGTATGGCCATTAACGTTGGCATAGATAAG	367
DB	305	GAATCAAAAACGTCACAGCAGGTGCAATCCAATCCGTTATTCGTCGTGGGATTGAACAG	364
QY	368	CTGTTGTTGCTTACTTAAAGAACTAAGCGACATTAACAAAGCCCTACTCGTGAACAAAG	427
DB	365	CAGTTGCCGAGCGTTGAAGCTTTGAAACCAACGCTATCCCTGTTGCCAATTAAGAG	424
QY	428	AAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAAATAGGTAATATCATAG	487
DB	425	CTATCGCTCAAGTTGCGACCGTATCTTCTGCTCTGA--AAAGTTGGTCACTACATCT	481
QY	488	CTGAAGCTATGGCTAAAGTTGGAAGAGGAGTGTATCAGAGTTGGAAGAGCTAAAGTC	547
DB	482	CTGAAGCAATGGAAGAGGTTGGCAAGACGCTGTGATCACCATCAAGAGTCAAGTGGTA	541
QY	548	TTGAACTACATTAAGTGTGTTGAAGAAATGAAGTTGACCGTGGCTACTCTCTCAT	607
DB	542	TGGAACAGAGCTTGAAGTGTGTAAGGAATGCAAGTTGACCGTGGTATACCTTTACAGT	601
QY	608	ACTTTGTAACCTAATCCCTGAGAAATGTTTGTGAACTTGTAACTTGTAACTTATATCTTTGTA	667

RESULT 6

US-08-961-527-77
; Sequence 77, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch

DB	602	ACATGCTGACAGATAGCGAAAAAATGCTGTGACCTTGAATAATCCGTACATTTTGATTA	661
QY	668	ATGAGAAAAGATTTACTAGCATGAACACATGCTTACCAATCTTTAGACAAGATCTCTAAG	727
DB	662	CAGACAAGAAAAATTTCCAATATCAAGAAATCTTGCCACATTTTGGAAGCAATCTTCCAAA	721
QY	728	TAAACGCTCCACTCTCTTTATTTGCTGAAGACGCTAGAGGTGAAGCACTTTGCAACACTTG	787
DB	722	GCATCGTCCACTCTTTGATTTATGCGGATGATGATGATGATGATGATGATGATGATGATG	781
QY	788	TAGTCAATAAGCTCCGCTGAGACCTCAAGTTGTCGCGTAAAGCTCTCTGTTTGGT	847
DB	782	TTTTGAACAAGATTCGTGGAACCTTCAACGCTAGTAGCAGTCAAGGCACTCTGTTTGGT	841
QY	848	AAGCCGTTAAGCTATGCTTTGAAGATTTGCTATCTTACTGAGGAGAGACATATTTG	907
DB	842	ACGCTGCAAGGCAATGCTTTGAAGATATCCCATCTTTAAGCGGCAACAGTTATCAAG	901
QY	908	AAGATCGTGTATATAAGCTTTGAAATATGAAGCTTTGCTCTTTTGAAGCAAGCTTAAAGCTG	967
DB	902	AAGACCTTTGCTCTTGAAGATGCGACAATTTGAAGCTCTTGGTCAAGCAGGAGAG	961
QY	968	TAGTTATGCAAGAAAAATATCTATCTGTTGATGCTGCTGGAATATCAGAAATATTA	1027
DB	962	TGACCGTGGACAAAAGATAGCAGCTTATTTAGAAAGGTGAGGAAATCTCTGAAGCGATTT	1021
QY	1028	AAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAACTGAAGCTCAGATATGATCGT	1087
DB	1022	CTCACCGTGTGGGTTATCAAGTCTCAAAATCGAAACTTCAACTCTTGAATTTGACCGTG	1081
QY	1088	AAAAAATTCAGAAACGCTTTGCAAACTCTTGTGTGAGTAGCTGTTTATCCATCTCGAG	1147
DB	1082	AAAAATTCAGAAACGCTTTGCAAACTCTTGTGTGAGTAGCTGTTTATCCATCTCGAG	1141
QY	1148	CTGCTACTGAACTGAATGAAAG	1207
DB	1142	CCGCAACTGAACTGAGTTGAAAGAAATGAACTCCGCAATTTGAAGATGCGCTCAACGCTA	1201
QY	1208	CAAGAGCTCGGTTGAAGAGGATTTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1267
DB	1202	CTGCTGAGCTGTTGAAGAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1261
QY	1268	TGAAGTCTTGTGATGATTAATAAAGCTCTGCTGATGATGATGATGATGATGATGATGATGATG	1327
DB	1262	TTCCAGCTGTTGCTACCTTGAATTTGACAGAGATGAGCA--ACAGAGCTATATTTG	1318
QY	1328	TCCGTCCTTCTTTGAAGAGCTTTTACGTCAAAATGCTGCAAAATGCTGCTGCTGCTGCTGCTG	1387
DB	1319	TTCTCCGCTGCTTTGGAAGAACCTTGTGCTCAAAATTTGCTCAAAATGCTGCTGCTGCTGCTG	1378
QY	1388	CTATTGTTGTAAGAAAGTTGTAACCAAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1447
DB	1379	CTATCGTATTCGATCGCTTTGAAAAATGCTGAGCTTGTATAGGATTTCAACGCGAGCACTG	1438
QY	1448	GAGAATATGAAGACCTTTATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1507
DB	1439	GCGAGTGGGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1498
QY	1508	CATTACAAAATGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1567
DB	1499	CCCTACAAAATGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1558
QY	1568	AAAAACAGAAC	1580
DB	1559	ATAAACAGAAC	1571

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 77:

SEQUENCE CHARACTERISTICS:

LENGTH: 5365 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-77

Query Match 41.7%; Score 687.6; DB 4; Length 5365;
Best Local Similarity 65.7%; Pred. No. 4.8e-171;
Matches 1034; Conservative 0; Mismatches 534; Indels 6; Gaps 2;

QY 7 TCTAAGAAATCCTTTTGGATGCTAAAGCCCGTGAAAACTTTCAGAGGTGTAGATAAA 66
DB 281 TCAAAGAAATTAATTTTCATCAGATGCCCGTTCAGCCATCGTTTCGTGGTGTGATATC 340
QY 67 CTTGCAAAATCGTTAAAGTAACACTTGGACCTTAAAGCGGTAATGTCGTATTTGAAAG 126
DB 341 CTTGCAACACCTGTTAAAGTAACCTTGGACCAAAAGTCCGATGTCGTTCTTGAAG 400
QY 127 TCTTTTGGTCCCGAGTATTACAAAGATGCTGTATCTGTTGCAAAAGAAATTTGAAC 186
DB 401 TCATTTCGGTTCACCCCTTGATTACCAATGACGGTGTGACCATTTGCCAAAGAAATTCGA 460
QY 187 GAAGATAAGTTTGAATAATGGCGCTCAATGGTTAAAGAGTAGTCTCCAAAAGTAC 246
DB 461 GAAGACCAATTTTGAATAATGGCGTCAATGGTTAAAGAGTAGTCTCCAAAAGTAC 520
QY 247 GATATTGCTGGTGGTAACTACACAGCAACAGTCTTTCGCAACAGTATTTATCGTGA 306
DB 521 GATATCGCAGGTGAGGAACTAGACTGCAACAGCTTTCGACCAAGTATCGTCCGTA 580
QY 307 GGTGTAAACCTTGTAGCAGCTGCTGTAATCTCTATGGCCATTTAAAGCTGGCAGTAA 366
DB 581 GGAATCAAAACGTACACAGGTGCAATCCAATCCAATCGGTATTCGTGCGGATTTGAACA 640
QY 367 GCTGTTGTTGCTTACTAAAGACTAAGCGCATTTAAGAGCTACTCGTGACCAAAA 426
DB 641 GCAAGTTGCGCGAGCTTGAAGCTTTGAAAAACAACGCCATCCCTGTTGCCAATAAGAA 700
QY 427 GAAATAGCTCAAGTTTGAAGCACTTCTGCAAACTCTGTATACCAACATAGTAAATATA 486
DB 701 GCTATCGCTCAAGTTGACCGGTATCTCTGTTCTGA---AAAAGTTGGTGAATATC 757
QY 487 GCTGAAGCTATGGCTAAAGTTGAAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAG 546

DB 758 TCTGAAGCAATGCAAAAAAGTTGCAAAAGACGGTGTATCACCATCGAAGAGTCACTGCT 817
QY 547 CTTGAAACTACATTAGATGTGTTGAAGGAATGAAGTTGACCGTGGCTACCTCTCTCCA 606
DB 818 ATGGAACACAGAGCTTGAAGTGTGAAGGAATGCAAGTTGACCGTGGTATACCTTTCA 877
QY 607 TACTTTGTAACCTAATCCTGAGAAATGGTTGTAACCTTGAATGTAACCTTATATCTTT 666
DB 878 TACATGGTGACAGATAGCGAAAAATGGTGGCTGACCTTGAATACTCGTACATTTGAT 937
QY 667 AATGAAAAAGATTACTAGCATGAAAGACATGCTTCAATCTTGAACAAGTGTGATAA 726
DB 938 ACAGACAAGAAAAATTTCCAATATCAAGAAATCTTGCCACTTTTGGAAAGCATTTCT 997
QY 727 GTAAACCGTCCACTCTTATATTTCTGTAAGAGTAGAAGGTGAAGCACTTCCACACT 786
DB 998 AGCAATCGTCCACTCTTATTTATCGGATGATGGATGGGAGGCTCTTCCAACTCT 1057
QY 787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAAGTTGTAGCCGTAAAAAGCTCTCTGTT 846
DB 1058 GTTTTGAACAGATTCTGTGAACCTTCAACGTAGTAGCAGTCAAGSCACCTGGTTTGT 1117
QY 847 GAACCGGTAAAGCTATGCTTGAAGATATGCTATCTTCTTCTTTAGGAACAGCTAAAC 966
DB 1118 GACCGTCCAAAGCCATGCTTGAAGATATCGCCATCTTAAACAGCGCAACAGTTATCA 1177
QY 907 GAAGATCGTGGTATAAAGCTTGAAATGTAAAGTGTCTTCTTTAGGAACAGCTAAAC 966
DB 1178 GAAGACCTTGGTCTTGAAGTGAAGATGCAATTTGAAGCTCTTGGTCAAGCAGGAGA 1237
QY 967 GTAGTTATTGACAAAGAAATACTACTATCGTTGATGTTGCTGGAATACTCAGAA 1026
DB 1238 GTGACCGTGGCAAGATAGCACGGTATTGTAGAGGTGCAAGGAAATCCTGAAGCAT 1297
QY 1027 AAAGCTCGAGTTAAACAAATTCGTCACAAATTTGAAGAAACAAAGCTCAGATTATGAT 1086
DB 1298 TCTCACCCTGTTGCGGTTATCAAGTCTCAAACTCGAAATCGAACTTCTGAATTTGACC 1357
QY 1087 GAAAACTTCAGAAAGCTTTCAAAACCTTGTGTTGGAGTAGCTGTATCCATGTTGA 1146
DB 1358 GAAAAATTTGCAAGAACGCTTGGCCAAATTTGTCAGGTGGTGTAGCGGTATTAAGGTT 1417
QY 1147 GCTGCTACTGAACTGAAATGAAAGAAAGAGATCGTGTAGAGATGCTCTTAAATGCA 1206
DB 1418 GCGCAACTGAACTGAGTTGAAGAAATGAAACTCCGATTTGAAGATGCCCTCAAGCT 1477
QY 1207 ACAAGAGTCCGGTTGAAGAGGATATGTCCTGTTGGTGGTGTACTGCTTTTGTCCGCT 1266
DB 1478 ACTCGTGCAGCTGTTGAAGAGGATATGTTGCAAGGTGGTGGAAACAGCTCTTGC 1537
QY 1267 ATTAAAGCTCTTGATGATATTAACCTCTGATGATGATGAACCTCTGAGCTTAATATC 1326
DB 1538 ATTCAAGCTGTTGCTTACCTTGGAAATTCAGAGAGATGAAGCA---ACAGCGTAA 1594
QY 1327 ATCCCTGCTCTCTTGAAGAGCTTTTACGTCAAAATTTGCTGCAAAATGCTGGCTAT 1386
DB 1595 GTTCTCCGTTGTTGAAGAACCCGTTGCTCAAAATGCTCAATGCAAGATTTGAAGA 1654
QY 1387 TCTATTGTTGTAAGAAAGTTCGTGAACCAAAAGATGGTTTGGATTTTAATGCTGATCA 1446
DB 1655 TCTATCGTTATCGATCGTTTGAATAATGCTGAGCTTGGTATAGGATTTAAACGCA 1714
QY 1447 GGAGATATGAAGCTTTATTAAGCTGTGTGCTTGTATGATCTTAAAGAGTTACAGTAT 1506
DB 1715 GCGAGTGGGTTAAATGATGATCAAGGATCATTTGATCCAGTTAAAGTGTAGTCT 1774
QY 1507 GCATTAACAAATGCAAGCTAGTCTCTTCTTAACTTAACTACAGAAATCGCTATGCT 1566
DB 1775 GCGCTTACAAATGCAAGCTAGTCTTAGCCAGCTTGAATTTTGACAAACAGAGCT 1834
QY 1567 GAAAAACCAAGAAC 1580

Db 1835 AATAAACCCAGAAC 1848

RESULT 7

US-09-472-971-5

; Sequence 5, Application US/09472971

; Patent No. 6197547

; GENERAL INFORMATION:

; APPLICANT: SOGO, Kazuo

; APPLICANT: YANAGI, Hideki

; APPLICANT: YURA, Takashi

; TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS

; FILE REFERENCE: 1422-409P

; CURRENT APPLICATION NUMBER: US/09/472,971

; CURRENT FILING DATE: 1999-12-28

; EARLIER APPLICATION NUMBER: JP10-372965

; EARLIER FILING DATE: 1998-12-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1647

; TYPE: DNA

; ORGANISM: Escherichia coli

; US-09-472-971-5

Query Match 41.5%; Score 682.8; DB 3; Length 1647;

Best Local Similarity 63.4%; Pred. No. 5e-170;

Matches 1044; Conservative 0; Mismatches 602; Indels 0; Gaps 0;

QY 1 ATGGCTTCTAAAGAAATCTTTTGTGCTAAAGCCGCTGAAAGAACTTTTCCAGAGGTGTA 60

Db 1 ATGGCAGCTAAAGACGTAAATTCGGTAACGAGCTCGTGTGAAATGCTGCGCGGCTA 60

QY 61 GATAAATCTCAATGCTGTTAAAGTAACACTTGGACCTAAAGCCGCTAATGCTTAAT 120

Db 61 AAGCTACTGCGAGATGAGTGAAGTGTACCTCGTCCAAAGCCGCTAAGTAGTTCG 120

QY 121 GAAAGTCTTTTGGTCCCGCAGTATTACAAAGAGATGCTGTAATCTGTTGCAAGAAAT 180

Db 121 GATAAATCTTTTGGTCCCGCAGCATTACCAAGAGATGCTGTTCCGTTGCTCGTGAATC 180

QY 181 GAACTTGAAGATAGTTTGAATAATATGGCGCTCAAAATGTTAAAGAACTAGCTCCCAAA 240

Db 181 GAACTGGAAGACAGTTCGAATAATATGGTGGCAGATGTTGAAGAACTGCTCTAAA 240

QY 241 ACTAGAGATATGCTGATGAGAACTAACAGCAACAGTCTTTCGCAAGCTATTTAT 300

Db 241 GCAAAACGACCTGCGAGCGACGGTACCACTGCAACCGTACTGCTCAGGCTATCATC 300

QY 301 CGTGAAGGTGTAATACTTGTAGCAGCTGGTCTGTAATCTTATGCGCATTAACCTGGCATA 360

Db 301 ACTGAAGGTGTAATACTTGTGCGGGCATGAACCCGATGACCTGAAACCGTGGTATC 360

QY 361 GATAAAGCTGTTGTTGTTTACTTAAAGAACTAAGCGACATTAACAGCCCTACTCGTGAC 420

Db 361 GACAAAGCGTTTACCCTGCGAGTTGAAGAACTGAAGCGCTGTCCTGACCTCTGAC 420

QY 421 CAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATCAACAAATAGGTAAT 480

Db 421 TCTAAAGCGATTGCTCAGGTTGTTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 480

QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAGAGAGGTGTTATCATCAGTTGAGGAAGCT 540

Db 481 CTGATCGCTGAAGCGATGGAACAAAGTCGGTAAAGAGCGGTATATCACCGTTGAAGACGGT 540

QY 541 AAAAGGTCTTGAAACTCATATTAGATGTTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600

Db 541 ACCGGTCTGAGGACGAACTGGACGTGTTGTTGAAGTATGCAAGTTCGACCGTGGCTACCTG 600

QY 601 TCTCCATCTTTTGAATTAATCTCGAAGAAATAGTTTGTGAACCTTGATTAACCTTTATTC 660

Db 601 TCTCCTTACTTTCATCAACAAGCCGAAACTGGCGCAGTAGAACTGGAAGACCCCGTTCATC 660

QY 661 CTTTGTAAATCAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTTAGAACAAAGTT 720

Db 661 CTGCTGCTGACAGAAAAATCTCCACATCCGCGAAATGCTGCGGTTCTGGAAGCTGTT 720

QY 721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGCTAGAAGGTGAAGCACTTGCA 780

Db 721 GCCAAAGCAGGCAAAACCGCTGCTGATCATCGTGAAGATGTAGAAGCGGAAGCGCTGGCA 780

QY 781 ACATTGTAGTCAATAAGCTCCGTTGAGAGCACTCAAGTTGAGCCGTAAGAGCTCTGGT 840

Db 781 ACTGCTGTTGTTAAACCACTTCGTTGAGCACTGTAAGATCGTGGTGAAGCAACCGGC 840

QY 841 TTTGTTGAACCGCTAAAGCTATGCTTGAAGATATTCCTTACTTACTGAGAGAGAGCA 900

Db 841 TTGCGCATCTGCTAAAGCTATGCTCCAGATATCGCAACCCCTGACTTGGCGGTACCGTG 900

QY 901 ATATTTGAAGATCGTGTATATAAGCTTGAAGCTTGAAGCTTGTCTTCTTTAGGAACAGCT 960

Db 901 ATCTCTGAAGATCGTGTATGGAGCTGGAAGAAAGCAACCTGGAGACCTGGGTGAGCT 960

QY 961 AAAGCTGTAGTATTGACAAAGAAATACTACTATCTGTTGATGCTGCTGAGAAATCAAG 1020

Db 961 AAAGCTGTTGTTGATCAACAAAGACACCCACTATCATCGATGCGGTGGGTGAAGAGCT 1020

QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAAGAAACCAAGCTCAGATTAT 1080

Db 1021 GCAATCAGGCGCGTGTGCTGATCCGTCAGATCCGTCAGAGATGGAAGAGCACTTCTGACTAC 1080

QY 1081 GATCGTGAAGAACTTCAAGAACTCTTTGCAAACTTTGTTGGTGGAGTGTGTTATCCAT 1140

Db 1081 GACCGTGAAGAACTGTCAGGAAACCGTAGCGAAACTGSCAGCGCGTTCAGTATCAAA 1140

QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAGAAAGAGAGATCGTGTGAAGATGCTCTA 1200

Db 1141 GTGGGTGCTCTACCGAAGTTGAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

QY 1201 AATGCAACAGAGCTGCGTTGAAGAGGATTTGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

Db 1201 CAGCGACCGCTGCTGCGGTGAGAGAGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

QY 1261 CGCTCCATTAAGCTCTTGTATGATTAATAACCTGCTGATGATGAATTAATTTGCTGAATTT 1320

Db 1261 CGGCTAGCGCTAAACTGCGCTGACCTCGCTGCTCAGAACGAGACAGACGCTGGGTATC 1320

QY 1321 AATATCATCGTCTGCTTCTTGAAGAGCTTTTACGTCAAAATGCTGCAAAATGCTGGCTAT 1380

Db 1321 AAAAGTTGCACTGCGTGCATGAAGAGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380

QY 1381 GAAGGTCTTATTTCTGTAGAAAAGTTTCGTGAACCAAAAGATGGTTTGTGATTTAATGCT 1440

Db 1381 GAACCGTCTGTTGTTGCTAAACCGTTAAAGCGCGGACCGCAACTACCGTTTACACGCA 1440

QY 1441 GCATCAGGAGATATGAAGACCTTATTAAAGCTGGTGTCTATTCATCCCTAAAAAGTTTACA 1500

Db 1441 GCAACCGAGAAATACCGCAACATGATCGACATCGGTATCTCTGATCAACCAAGTAAT 1500

QY 1501 CGTATTGCTATTAATAATGAGCATCAGTACCTCTTACTTTCTAATCAGAAATGCGCT 1560

Db 1501 CGTTCTGCTCTGAGTACGAGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

QY 1561 ATTGCTGAAAACCAAGACCTTAAAGAGATATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTG 1620

Db 1561 GTTACCGACCTGCGGAAACCGATGACGCTGACTTAGGCGCTGCTGCGGTATGGGCGGC 1620

QY 1621 ATGGGTGCTATGACCGTATGTACTA 1646

Db 1621 ATGGGTGCTATGCGCGCATGTGTA 1646

RESULT 8

US-09-472-971-7

; Sequence 7, Application US/09472971

; Patent No. 6197547

GENERAL INFORMATION:

APPLICANT: SOGO, Kazuyo
 APPLICANT: YANAGI, Hideki
 APPLICANT: YURA, Takashi
 TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
 FILE REFERENCE: 1422-409P
 CURRENT APPLICATION NUMBER: US/09/472,971
 EARLIER FILING DATE: 1999-12-28
 EARLIER APPLICATION NUMBER: JP10-372965
 EARLIER FILING DATE: 1998-12-28
 NUMBER OF SEQ ID NOS: 7
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 7
 LENGTH: 4524
 TYPE: DNA
 ORGANISM: Escherichia coli
 US-09-472-971-7

Query Match 41.5%; Score 682.8; DB 3; Length 4524;
 Best Local Similarity 63.4%; Pred. No. 8.1e-170;
 Matches 1044; Conservative 0; Mismatches 602; Indels 0; Gaps 0;

QY 1 ATGGCTTCTAAGAAATCCCTTTTGTGATGCTAAGCCCGTGAAGAACTTTTCAGAGGTGA 60
 DB 383 ATGGCAGCTTAAGAGCTAAATTCGGTAAACGACGCTCGTGTGAATAATGCTGCGCGCGTA 442
 QY 61 GATAAAGTGCAGAAATGCTGTTAAAGTAACTTGAAGTAAAGGCGGTAAATGCTGCTTAT 120
 DB 443 AAGCTACTGCGCAGATGCGATGAAGTTACCTTCGGTCCAAAGGCGGTAAAGTACTG 502
 QY 121 GAAAAAGTCTTTTGGTTCCTCCAGTATTAACAAAGATGCTGATCTGTTGCAAAAGAAAT 180
 DB 503 GATAAATCTTTCGGTGCACCGACCATCACCAAGATGCTGTTCCGTTCTCGTGAATC 562
 QY 181 GAACCTGAAGTAAGTTGAAATATGCGCGCTCAAAATGCTGTAAGAGTACTGCCAAA 240
 DB 563 GAACCTGAAGTAAGTTGAAATATGCGCGCTCAAAATGCTGTAAGAGTACTGCCAAA 622
 QY 241 ACTAGCGATGCTGCTGATGAACTCAACAGCAAGTCTCTGCAAGCTATTTAT 300
 DB 623 GAAACGAGCTGCGAGGCGACGCTACCACTGCAACCGTACTGCTCAGGCTATCATC 682
 QY 301 CCGTGAAGTGTAAACCTTCTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 DB 683 ACTGAGGCTCTGAAGCTTCTGCTGCGGCATGAACTGCAAGCTGCAAGCTGCAAGCT 742
 QY 361 GATAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 DB 743 GACAAAGCGGTTACCGCTGAGTGAAGAACTGAAAGCGCTGCTGCTGCTGCTGCTGCTG 802
 QY 421 CAAAGAAATAGCTCAAGTTCGAACCAATTCGCAAACTCTGATACCAATAGTAAAT 480
 DB 803 TCTAAGCGATGCTCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 862
 QY 481 ATCATAGCTGAAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 DB 863 CTGATCGCTGAAGCGATGACAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 922
 QY 541 AAGGCTCTGCAACATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 923 ACCGCTCTGCAAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 982
 QY 601 TCTCCATCTTTGTAATCTCTGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 983 TCTCTTCT 1042
 QY 661 CTTTGTAAAGTGAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 DB 1043 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1102
 QY 721 GCTAAAGTAAACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 DB 1103 GCCAAAGCGGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1162

QY 781 ACACCTTGCTAGTCAATAGCTCCGTGAGCACTCCAAAGTTGTAGCCGTAAGCTCTCTGCT 840
 DB 1163 ACTGCTGTTGTTAAACCACTTCTGTCGATCGTGAAGTCTGCTGCTGCTGCTGCTGCTGCT 1222
 QY 841 TTTGCTGAAGCGCGTAAAGCTATGCTGGAAGTATGCTATCTTCTGCTGCTGCTGCTGCTGCT 900
 DB 1223 TTTGCGGATCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1282
 QY 901 ATATTGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 DB 1283 ATCTCTGAAGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1342
 QY 961 AAACTGCTAGTCTTATGCAAAAGAAATATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 DB 1343 AAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1402
 QY 1021 GATATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 DB 1403 GCAATCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1462
 QY 1081 GATCTGGAAGAACTTCAAGACGCTCTGCAAACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 DB 1463 GACCTGGAAGAACTTCAAGACGCTCTGCAAACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1522
 QY 1141 GTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 DB 1523 GTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1582
 QY 1201 AATGCAACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 DB 1583 CACGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1642
 QY 1261 CGCTCAATTAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 DB 1643 CCGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1702
 QY 1321 AATATCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 DB 1703 AAGTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1762
 QY 1381 GAAGGTTCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 DB 1763 GAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1822
 QY 1441 GCATCAGGAGATATGAAGACCTTATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 DB 1823 GCAACCGAAGATACCGCAACATGATCGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1882
 QY 1501 CGTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 DB 1883 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1942
 QY 1561 ATTGCTGAAACCGCAACCTTAAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 DB 1943 GTTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2002
 QY 1621 ATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1646
 DB 2003 ATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2028

RESULT 9

US-08-470-260-7
 ; Sequence 7, Application US/08470260
 ; Patent No. 607706
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; APPLICANT: Bugnoli, Massimo
 ; APPLICANT: Telford, John
 ; APPLICANT: Macchia, Giovanni
 ; APPLICANT: Rappuoli, Rino
 ; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful

TITLE OF INVENTION: for Vaccines and Diagnostics
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/470,260

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/256,848

FILING DATE: 21-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0316.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2708

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1838 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-470-260-7

Query Match 39.08; Score 642.4; DB 3; Length 1838;

Best Local Similarity 62.98; Pred. No. 2.26-159;

Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;

QY 8 CTAAGAAATCCCTTTTGTGCTAAAGCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
DB 62 CAAAGAAATCAATTTTCAGATAGTCGAGAACTTTTATTGAAGCGGTGAGGCAAC 121
QY 68 TTGCAATGCTGTAAAGTAACCTTGGAACCTAAGGCGGTAAAGTCTGTTATTGAAGAAGT 127
DB 122 TCCATGACGCTGTCAAAGTAACCATGGGCGCAGAGGCGAGATGTATTGATCCAAAAA 181
QY 128 CTTTGTGTTCCCGAGTTATTACAAAGATGGTGTATCTGTGCAAAAGAAATGAACCTTG 187
DB 182 GCTATGCGCTCCCAAGCATCAACAAAGCGCGTGGCTGAGAGAGATTGAATTA 241
QY 188 AAGATAGTTTGAATAATGGGCTCAAAATGTTAAAGAGTAGCTCCCAAACTAGCG 247
DB 242 GTTGCCAGTAGCTAAACATGGGCTCACTCTGTTAAGAGTAGCGAGCAAAACCGCTG 301
QY 248 ATATTGCTGTGATGGAACCTAACACAGCAAGCTTGTGCAAGCTTATTATCGTGAAG 307
DB 302 ATGCTGCGCGGATGGCAGCAGCAGCAGCGCTGCTAGCTTATAGCATTTTAAAGAG 361
QY 308 GTGTAAACTGTAGAGCTGGTCTGTAATCTATGGCCATTAAACGTTGGCATAGTAAG 367
DB 362 GTTTGAGGAATATACCGGCTGGGCTAACCCCTATTGAAAGTGAACGAGGCATGATAAG 421
QY 368 CTGTTGTGCTTACTAAAGAACTAAGCGACATTACAAAGCCCTACTCGTCAACCAAAAG 427
DB 422 CTGCTGAAGCGATCATTTATGAGCTTAAAGAGCGCAAAAAGTAGGCGGTAAAGAG 481
QY 428 AAATAGCTCAAGTTGGAACCACTTTCTGCAACTCTGTACAACTAGTAATATCATAG 487
DB 482 AAATCAACCCAGTGGGACCACTTTCTGCAAACTCCGATCAATATCGGGAACCTACG 541

QY 488 CTGAAGCTATGGCTAAAGTTGAAAAAGAGGTGTATTACAGTTGAGGAAGCTAAAGGTC 547
DB 542 CTGACGCTATGGAAGAAAGTGGGTAAGACGCGGTGATCACCGTTGAGGAAGCTAAAGGCA 601
QY 548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACTCTCTCCAT 607
DB 602 TTGAAGATGAATTTGGATGTCTGTAAGAGGCATGCAATTTGTATAGAGGCTACTCTCCCTT 661
QY 608 ACTTTGTAATCTCTGAGAAAAATGTTTGTGAACCTTGTAACCCCTTATATCTCTTGA 667
DB 662 ATTTTGTAGGAACGCTGAGAAATGACCGCTCAATTTGGAATATGCTTATCATCTTTTAA 721
QY 668 ATCAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTTAGAACAAAGTTGCTAAAG 727
DB 722 CGGATAAAAAATCTCTAGCATGAAGACATTTCTCCGCTACTAGAAAAAACCACTGAAG 781
QY 728 TAAACCGTCCACTCTCTTATTATTGCTGAAGACGTAGAGTGAAGCACTTGCACACTTG 787
DB 782 AGGCAAAACCGCTTTTAAATCATCGCTGAAGACATTTGAGGCGGAAGCTTTAACGACTTAG 841
QY 788 TAGTCAATAGCTCCGTGGAGCACTCCAGTCTGAGCGTAAAGCTCCCTGTTTTCGTG 847
DB 842 TGGTGAATAAATTAAGAGCGGTGTTGAATTCGAGCGGTTAAAGCTCCAGGCTTTTGGG 901
QY 848 AACGCGCTAAAGCTATGCTTTGAAGATATTGCTATCTCTTACTGGAGGAGAAAGCAATATTG 907
DB 902 ACAGAGAAAGAAATGCTCAAGACATCGCTATTTAACCGCGGTCAAGTCATTAGCG 961
QY 908 AGATCGTGTATAAAGCTTTGAAATGTAAGCTTTGCTTTCTTTAGAACAGCTAAACGTG 967
DB 962 AAGAAATTTGGCTTTGAGTCTAGAAAAAGCTGAAGTGGAGTTTATTAGGCAAGCTGGAAG 1021
QY 968 TAGTTATTGCAAGAAATACTACTATCTGTCATGTCGTGCGTAAATACAGAGATTA 1027
DB 1022 TTGTGATTGCAAAAGACACACCAAGCATGTAGTGCAGAAAGCCATGAGCATGATGTTA 1081
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAAAGCTCAGATTATCATCGT 1087
DB 1082 AAGACAGCTCGCGAGATCAAAACCAATTCGAAGTACGACAGGATTTATGACAAAG 1141
QY 1088 AAAAATCTCAAGAACGCTTTTGCAAACTTTGTTGGTGAAGTAGTGTGTTATCCATTTGAG 1147
DB 1142 AAAAATTCAGAAAGATTTGGCTAAACTCTCTGCGGTGTGGCTGTGATTAAAGTGGCG 1201
QY 1148 CTGCTACTGAACTGAAATCAAAAGAGAGAGATCGTGTAGAGATGCTCTTAATCAAC 1207
DB 1202 CTGCGAGTGAAGTGAATCAAAAGAGAAAGACCGGTTGGATGACCGCTTTGAGCGCGA 1261
QY 1208 CAAGAGCTGCGGTTGAAGAAAGTATTGTCCTGTCGTGTCGTCTGCTTTTGTCCGCTCCA 1267
DB 1262 CTAAGCGCGGTTGAAGAGGCATTTGATTTGTCGCGGTGCGGCTCTCATTTCCGCGG 1321
QY 1268 TTAAGTCTTGTATGATATTAAACCTGCTGATGATGATGAACCTGCTGGACTTAAATCA 1327
DB 1322 CTCAAAAGTGCATTTCAATTTGC-----ACGATGATGAAAAAGTGGGCTATGAAATCA 1375
QY 1328 TCCGTCCTTCTCTGAGAGCCTTTACGTCAAAATGCTGCAAAATGCTGGCTATCAAGGTT 1387
DB 1376 TCATGCGCGCATTAAGCCCTTAGCTCAAAATGCTATCAACGCTGGTTATGATGGCG 1435
QY 1388 CTATTGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGTTTGGATTTTAATGCTCATCAG 1447
DB 1436 GTGTGTCGTGATGAAGTAGAAAAACACGAGGGCATTTTGGTTTAAACGCTAGCAATG 1495
QY 1448 GAGAAATGAAGACCTTTTAAAGCTGGTGTGATTCATTCCTTAAAAAAAGTTTACAGTATG 1507
DB 1496 GCAAGTATGTGGATATGTTTAAAGAGGCATTTATTGACCCCTTAAAAAGTAGAAGGATCG 1555
QY 1508 CATTACAAATGACGATCAGTAGCCCTCTTACTTCTAACTACAGATGCGCTATTCTG 1567
DB 1556 CTCTACAAATGCGGTTTCGTTTCAAGCCCTGCTTTTAAACACAGAGCCACCGTGCATG 1615
QY 1568 AAAAACCAAGAACCTTAAAAAAGATATGCTATGCTG---GCGGTGTTATGGGTGCTATG 1624

Db 1616 AAATCAAGAGAGAAAGCGACTCCGGCAATGCTGATATGGTGGCATGGCGGTATGG 1675
QY 1625 GTGGTATGACGGTATGTACTA 1646
Db 1676 GAGGATGGCGGCATGATGTA 1697

RESULT 10

US-08-471-491-7
; Sequence 7, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-08-471-491-7

Query Match 39.0%; Score 642.4; DB 3; Length 1838;
Best Local Similarity 62.9%; Pred. No. 2.2e-159;
Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;
QY 8 CTAAGAATCCTTTTGTATGCTAAGCCCGTGAAGAACTTTCACGAGGTGTAGATAAC 67
Db 62 CAAGAAGAATCAAAATTTTCAGATAGTGCAGAGAAACCTTTTATTTGAAGCGGTGAGCAAC 121
QY 68 TTGCAAAATGCTGTTAAAGTAACACTTGGACCTAAAGCGCGTAACTGCTTATTTGAAAGT 127
Db 122 TCCATGACGCTGCTAAAGTAACCTGAGGCGCAAGGCGAGGAGTATGATTCGCAAAA 181
QY 128 CTTTGGTTCCTCCAGTTATTAACAAAGATGCTGATCTCTGTCGCAAGAAATTTGAACCTG 187
Db 182 GCTATGGCGCTCCAGCATCACCAAGACGGCGTGCGTGCCTAAGAGATTGATTA 241
QY 188 AAGATAAGTTTGAATAATGCGCGCTCAAAATGTTAAGAGTAGTCTCCCAAACTAGCG 247
Db 242 GTTGCCCAAGTACTAACATGGCGCTCAACTCGTTTAAAGAGTAGCGAGCAAAACCGCTG 301
QY 248 ATATTGCTGGTATGAACTACACAGCAACAGTCTTGCACAAAGCTATTTATCGTGA 307
Db 302 ATGCTGCGCGGATGGCGACACAGCGACCGCTGCTAGCTTATAGCATTTTAAAGAG 361
QY 308 GTGTAAACTTTGTAGCAGTGGTCTGCTAATCTTATGCCATTAAGCGTGGCATAGATAAG 367
Db 362 GTTTCAGGAATATCACGGCTGGGCTAACCCATTGAACTGAAACGAGGCATGGATAAG 421
QY 368 CTGTTGTTGCTTTACTAAGAACTAAGCGCATTAAGCGCTTACAAAGCGTCTCTGTGACCAAAAG 427
Db 422 CTGCTGAAGCGCATCAATTAATGAGCTTAAAGAAAGCGAGCAAAAGAGTAGCGGTAAAGAG 481
QY 428 AAATAGCTCAAGTTCGAAACCTTTCTGCAAACTCTGTATACACAAATAGTATATCATAG 487
Db 482 AAATACCAAGTGGGCACTTTCTGCAAACTCTGCAAACTCTGCAAACTCTGCAAACTCTG 541
QY 488 CTGAGCTATGGCTAAGAGTTGAAAGAGGAGTGTATCACAGTTCAGGAAAGCTAAAGGTC 547
Db 542 CTGACGCTATGAAAGAGTGGGTAAAGACGGGTGATCACCGTTTGGAGAGCTAAGGCA 601
QY 548 TTGAACCTACATTAGATGTTTGAAGAGTAAAGTTTGAACCGTGGCTACCTCTCCAT 607

Db 602 TTCAAGATCAATTTGGATGCTGTAGAGGCGATGCAATTTTGATAGAGGCTACCTCTCCCTT 661
QY 608 ACTTTCTACTAATCTGTAGAGAAATGGTTTGTCAACTTGTATACCTTATATCTCTTTGTA 667
Db 662 ATTTTGTAAAGAACGCTGTAGAGAAATGACGCTCAATTTGGATTAATGCTTATCATCTTTTAA 721
QY 668 ATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACAAAGTTGCTTAAAG 727
Db 722 CGGATAAAAAATCTCTAGCATGAAGACATTTCTCCGCTACTAGAAAAAACCATGAAG 781
QY 728 TAAACCGTCCACTCCCTTATTATTGTCTGAAGACGTGAAGGTGAAGCACTTGAACAACCTTG 787
Db 782 AGGGCAAAACCGCTTTTAAATCATCGCTGAAGACATTTGAGGCGCAAGCTTTAAACGACTCTAG 841
QY 788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTCTGTTTGGTG 847
Db 842 TGGTGAATAAAATTAAGAGCGGTGTTGAATATCGACGCGTTAAAGCTCCAGGCTTTGGGG 901
QY 848 AACGCGGTAAAGCTATGCTTTGAAGATATTGCTTACTCTGAGGAGAGCAATATTG 907
Db 902 ACAGAAGAAAGAAATGCTCAAGACATCGCTATTAAACCGCGCTCAAGTCAATTAGCG 961
QY 908 AAGATCGTGTATATAAGCTTTGAAATTTGAAGTGTCTTCTTTAGGAAACAGCTTAAACGCTG 967
Db 962 AAGAATTTGGCTTGAATCTAGAAAAACGCTGAAGTGGAGTTTATAGCAAAAGCTGGAAGGA 1021
QY 968 TAGTTATTGACAAAGAAATACTACTATCTGTTGATGCTGCTGGAATATCAGAGATATTA 1027
Db 1022 TTTGATTTGACAAAGACAAACACGATCGTAGATGGCAAGGCCATAGCGATGATGTTA 1081
QY 1028 AAGCTCGAGTTAAACAAATTTCTGTGCAAAATTTGAAGAAACAAAGCTCAGATTTATGATCGTG 1087
Db 1082 AAGACAGAGTCCGCGCAGATCAAAACCCAAATTTGCAAGTACGACAGCGATTAATGACAAAG 1141
QY 1088 AAAAATCTTCAAGACGCTTTGCAAAACCTTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db 1142 AAAAATTTGCAAGAAAGATTGGCTAAACTCTCTGGCGGTGCTGTGATTTAAAGTGGCG 1201
QY 1148 CTGCTACTGAACTCAATGAAAGAGAGAGAGATCGTGTAGAAATGCTCTTAATGCAA 1207
Db 1202 CTGCGAGTGAAGTGAAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261
QY 1208 CAAGAGCTCGCGTTGAAGAAAGTATTGCTCCCTGGTGGTGTACTGCTTTTGTCCGCTCCA 1267
Db 1262 CTAAGCGCGCTTGAAGAGGCAATTTGATTTGGTGGCGGTGCGCTCTCATTCGCGCG 1321
QY 1268 TTAAGTCTTGAATATTAACCTGCTGATGATGATGAACCTTCTGCTGACTTAATATCA 1327
Db 1322 CTCAAAAAGTGCATTTGAATTTGC-----ACGATGATGAAAAAGTGGGCTATGAAATCA 1375
QY 1328 TCCGTCGTTCTCTCAAGAGCGCTTTACGTCAAATTTGCTCAATGCTGCTATGAGGTT 1387
Db 1376 TCATCGCGCCATTAAGAGCCCATTTAGCTCAATCGCTATCAACGCTGGTTATGATGGCG 1435
QY 1388 CTAATTTGTTAGAAAAAGTTCTGTGAACCAAAAGATGGTTTGGATTTATGCTGCATCAG 1447
Db 1436 GTGTGCTGTGAATCAAGTAGAAGAAACAGAGAGGCAATTTGGTTTAAACGCTAGCAATG 1495
QY 1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTGCTGATGATCTTAAAGAGTTACAGCTATG 1507
Db 1496 GCAAGTATGTGATATGTTTAAAGAGGCAATTTTGAACCCCTTAAAGAGTAGAAGATCG 1555
QY 1508 CATTACAAAATGACATCAGTAGCTCTCTTACTTCTAAGTACAGAACTGCTATGCTG 1567
Db 1556 CTCTCAAAATGCGGTTTCGGTTTCAAGCTGCTTTTAAACCAAGCAACCGTGCATG 1615
QY 1568 AAAAACCCAGAACTTAAAAAGATATGCTATGCTG---GCGGTGATGGTGGTATGCG 1624
Db 1616 AAATCAAG 1675
QY 1625 GTGGTATGACGGTATGACTA 1646
Db 1676 GAGGATGGCGGCATGATGA 1697

RESULT 11

US-08-466-662-7
; Sequence 7, Application US/0846662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-08-466-662-7

Query Match 39.0%; Score 642.4; DB 3; Length 1838;
Best Local Similarity 62.9%; Pred. No. 2.2e-159;
Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;

QY	8	CTAAAGAAATCCCTTTTGATGCTAAAGCCGCTGAAAGAACTTTCACGAGGTGTAGATAAAC	67
DB	62	CAAAAGAAATCAAAATTTTCAGATAGTCGAGAAACCTTTTATTGGAAGCGTGAAGCAAC	121
QY	68	TTGCAATGCTGTTAAAGTAACACTTGGACCTTAAGCCGCTAAGTGTGTTATTGAAAGT	127
DB	122	TCCATGACGCTGTCAAAAGTAAACCATGCGGCAAGAGCGAGGAATGTATTGATCCAAAAA	181
QY	128	CTTTTGGTCTCCCAATATTACAAAGATGCTGTATCTGTGCAAAAGAAATGAACCTG	187
DB	182	GCTATGCGCTCCAAAGCATCACNAAGCGGTGAGCGGTGTTAAAGAGATTGAATTAA	241
QY	188	AAGATAAGTTGAAATATGCGGCTCAAAATGTTAAAGAAAGTAGCTCCCAAACTAGCG	247
DB	242	GTGCGCCAGTAGCTAAATGCGGCTCACTGTTAAAGAAAGTAGCGAGCAAAACCGCTG	301
QY	248	ATATTGCTGTGATGGAACCTAACACGACAGCTTCTTGCACAGCTATTATTGCTGAAG	307
DB	302	ATGCTGCGCGCATGGCAGCACACGACCGCTGTAGCTTATAGCATTTTAAAGAAAG	361
QY	308	GTCTAAAACTTGTAGCAGCTGCTGTAATCCTATGGCCATTAAACGCTGGCATAGATAAG	367
DB	362	GTTTGAGGAATATCACGGCTGGGCTTAACCTATTGAAAGTAAACGAGGCGATGATAAG	421
QY	368	CTGTTGTGCTGTACTTAAAGAACTAAGCGACATTAACAAAGCCCTACTCTGACCAAAAG	427
DB	422	CTGCTGAAGCGATCATTAATGAGCTTAAAGAAAGCGACAAAGTAGGCGGTAAAGAG	481
QY	428	AAATAGCTCAAGTTGAAACCATTTCTGCAACTCTGATACAACTAGGTAATATCATAG	487
DB	482	AAATACCCCAAGTGGCCACCATTTCTGCAAACTCCGATCAATATCGGGAACCTCATCG	541
QY	488	CTGAAGCTATGGCTAAAGTTGAAAGAGGCTGTATACAGTTGAGAGAGCTAAAGTTC	547
DB	542	CTGACGCTATGGAAGAGTGGTAAAGACGCGGTGATCACCGTTGAGAGAGCTAAAGGCA	601
QY	548	TTGAAACTACATAGATGTGTTGAAGCAATGAAGTTTGACCGGTGCTACTCTCTCCAT	607
DB	602	TTGAGATGAATTCGATGTCTGTAAGCGCATGCAATTTGATAGGGCTACTCTCCCTT	661
QY	608	ACTTTGTAACCTCTGAGAAATGTTTGTGAACCTTGATAACCCCTTATATCCTTTGTA	667
DB	662	ATTTTGAACGACGCTGAGAAATGACCGCTCAATTTGGATATGCTTACATCCTTTTAA	721

RESULT 12

US-08-256-847C-6
; Sequence 6, Application US/08256847C
; Patent No. 6403099

QY	668	ATGAGAAAAAGATTACTAGCATGAAGACATGCTACCAATCTTTAGAACAAAGTTGCTAAAG	727
DB	722	CGGATAAAAAAATCTTAGCATGAAGACATTTCCCGCTACTAGAAAAACCATGAAAG	781
QY	728	TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAGGTGAAGCACTTTGCAACACTTG	787
DB	782	AGGCAAAACCGCTTTTAAATCATCGCTGAAGACATTTAGGGCGAAGCTTTTAACGACTCTAG	841
QY	788	TAGTCAATAAGTCCGCTGGAGCACTCAAGTTGTAGCCGTAAGAGCTCTCGTGGTTGGTG	847
DB	842	TGTTGAATAAATTAAGAGCGGTGTTGAATATCGAGCGGTTAAAGCTCCAGGCTTTGGGG	901
QY	848	AAGCCCTTAAGCTATGCTTTGAAGATATGCTATCTTACTTGGAGGAGAGCAATATTG	907
DB	902	ACAGAGAAAGAAATGCTCAAGACATCGCTATTTTAAACCGCGCTCAAGTCAITAGCG	961
QY	908	AAGATCGTGTATATAAGCTTTGAAATGTAAGCTTTGCTTCTTTAGGAACAGCTAAAGCTG	967
DB	962	AAGAAATGGGCTTGAGTCTAGAAAACGCTGAAGTGGAGTTTATAGGCAAAAGCTGGAAG	1021
QY	968	TAGTTATTGACAAAATAACTACTATCGTTGATGCTGTGGAATAATCAGAAATATTA	1027
DB	1022	TTGTGATTGACAAAGACCAACACCGATCGTAGATGCAAAAGGCCATAGCGATGATGTA	1081
QY	1028	AAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGATCGT	1087
DB	1082	AGACAGAGTCGCGCAGATCAAAACCCAAATTCGAGTACGACNAGCGATTATGACAAAG	1141
QY	1088	AAAACTTCAAGAACGCTTTTGCAAACTTTGTTGTTGAGTAGCTGTTTATCCATGTTGGAG	1147
DB	1142	AAAAATGCAAGAAAGATTGGCTAAACTCTCTGCGGTGTGGCTGTGATTAAAGTGGCG	1201
QY	1148	CTGCTACTGAACTGAATGAAGAGAAAGAGATCGTTGTAAGAGTGTCTTAATGCAAA	1207
DB	1202	CTCGAGTGAAGTGAATGAAGAGAAAGAAACCGGGTGGATGACGCGTTGAGCGCGA	1261
QY	1208	CAAGAGCTGCGGTTGAAGAGAGTATGCTCCTGCTGTTGCTGCTGCTTTTGTCCGCTCCA	1267
DB	1262	CTAAGCGCGGTTGAAGAGGCAATTGTGATTGTTGCGGCTGCGGCTCTCANTTCGCGCG	1321
QY	1268	TTAAGTCTTGTGATGATTAATAAAGCTGCTGATGATGAATGAACTTGTGGAATTAATATCA	1327
DB	1322	CTCAAAAGTGCATTTGAATTTGC-----ACGATGATGAAGAGTGGGCTATGAAATCA	1375
QY	1328	TCCGTCTTCTTTGAAGAGCTTTAGTCAAAATGCTGCAAAATGCTGCTGATGAAGGTT	1387
DB	1376	TCATGCGCGCATTAAGACCCCATTTAGTCAAAATCGCTATCAACGCTGTTATGATGGCG	1435
QY	1388	CTATTGTTGTAGAAAAGTTTCGTGAACCAAAAGATGGTTTGGATTTAATGCTGCATCAG	1447
DB	1436	GTGTGTCGTGAATGAAGTAGAAAACACGAGAGGCAATTTTGGTTTAAACGCTAGCAATG	1495
QY	1448	GAGAATATGAAGACCTTTATTAAAGCTGGTGTCTATTGATCTCTAAAGAGTTTACAGTATTG	1507
DB	1496	GCAAGTATGAGATATGTTTAAAGAGGCAATTTTGAACCCCTTTAAAGTAGAAGAGATCG	1555
QY	1508	CATTACAAATGAGCATGAGTAGCCCTCTTACTTAAGTACAGAAATGCGCTATTGCTG	1567
DB	1556	CTCTACAAATGCGGTTTCGTTTCAAGCTGCTTTTAAACACAGAGACCGCTGCATG	1615
QY	1568	AAAAACGAGAACTAAAAGAGATATGCTATGCTG---CGGCTGTTGCTGCTGTTG	1624
DB	1616	AAATCAAGAGAAAGAGCGACTCCCGCAATGCTGATGTTGGTGGCATGGGCGGTATGG	1675
QY	1625	GTGCTATGACGCTATGTTACTA	1646
DB	1676	GAGGCAATGGCGCATGATGTA	1697

GENERAL INFORMATION:

APPLICANT: Rappuoli, Rino
APPLICANT: Costantino, Paolo
APPLICANT: No. 6403099elli, Francesco
TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
FILE REFERENCE: CHIR-0042
CURRENT APPLICATION NUMBER: US/08/256,847C
CURRENT FILING DATE: 1994-11-01
PRIOR FILING DATE: 1993-03-08
PRIOR APPLICATION NUMBER: PCT/EP93/00516
PRIOR APPLICATION NUMBER: F192A000058
PRIOR FILING DATE: 1992-03-06
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 1838
TYPE: DNA
ORGANISM: H. pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (58)..(1695)
OTHER INFORMATION:

US-08-256-847C-6

Query Match 39.0%; Score 642.4; DB 4; Length 1838;

Best Local Similarity 62.9%; Pred. No. 2.2e-159;

Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;

QY 8 CTAAGAAATCCTTTTGTGATGTAAGCCCGTGAATACTTTCACGAGGTGTAGATAAC 67
DB 62 CAAAGAAATCAAAATTTTCAGATAGTGCAGAAACCTTTTATTTGAAGCGGTGAGCAAC 121
QY 68 TTGCAATCTCTTAAAGTAACACTTGGACCTTAAGCCGTAATGTGTTTATTTGAAAAGT 127
DB 122 TCATGACCTGTCAAGATAACATGGGCGCAAGAGGAGGAGTGTATTTGATCCAAAAA 181
QY 128 CTTTGGTTCCTCCAGTTATTACAAAGATGTTGATCTCTTGCAGAAAGAAATTTGAACCTG 187
DB 182 GCTATGGCGCTCCAGCATCACCAAGAGCGGTGAGCGTGAAGATTTGAATTA 241
QY 188 AAGATAAGTTTGAATAATTTGGCGCTCAATGGTTAAAGATAGTCTCCAAATAGCG 247
DB 242 GTTGCCAGTAGTAACTAAGTGGCGCTCACTGCTTTAAAGAGTAGGAGCAAAACCGCTG 301
QY 248 ATATTGCTGTGATGAACTCAACAGCAACAGTCTTCCCAAGCTATTATTCGTGAAG 307
DB 302 ATGCTGCGCGGATGGCAGCACACAGCAGCGTGTAGCTATTAGCATTTTAAAGAG 361
QY 308 GTGTAACCTTTAGCAGCTGGTCTTAATCCTATGGCCATTAAACGTGGCATAGATAAG 367
DB 362 GTTTGAGGAATATACGGCTGGGCTAACCTTATTGAAGTGAACCGAGGCATGATAAG 421
QY 368 CTGTTGTTGCTTTACTAAGAACTAAGCGCATTAACAAGCCTACTCTGTACCAAAAG 427
DB 422 CTGCTGAAGCGATCATTTATAGCTTTAAAGAGCGAGCAAAAGTAGGCGGTAAAGAG 481
QY 428 AATAGCTCAAGTTGAACTTCTGCAACTCTGATACACAAATAGTGAATATCATAG 487
DB 482 AATACCCAGTGGCGACATTTCTGCAACTCCGATCACATATCTGGGAACTCATCG 541
QY 488 CTCAGACTATGGCTAAGTTGAAAGAGAGGTGTTATCACAGTTGAGGAAGCTTAAAGGTC 547
DB 542 CTGACGCTATGAAAGAGTGGTAAAGACGGGTGATACCGTTGAGGAAGCTTAAGGCA 601
QY 548 TTGAACACTAGATAGTTGTTGAAGGAATGAAGTTTACCGTGGCTACCTCTCTCCAT 607
DB 602 TTGAAGATCAATTTGATGCTAGAGGATGCAATTTTGTAGAGGCTACCTCTCCCTT 661
QY 608 ACTTGTAACTAATCTGAGAAATGTTTGTGAACCTTGATACCCCTTATATCTTTGA 667
DB 662 ATTITTAACGACGCTGAGAAATGACCGCTCAATTTGGATATGTTTACATCTTTTAA 721
QY 668 ATGAGAAAAAGATTACTAGCATGAAGAGATGCTACCAATCTTTAGAACAAAGTTGCTAAAG 727

DB 722 CGGATAAAAAAATCTCTAGCATGAAGACATTTCTCCGCTACTGAAAAAACCATGAAG 781
QY 728 TAAACCGTCCTCTCTTATTATTGCTGAAGAGCTAGAGGTGAACACTTGCACACTTG 787
DB 782 AGGCAAAACCGCTTTTAAATCATCGCTGAAGACATTTAGGCGCAAGCTTTAAACACTTAG 841
QY 788 TAGTCAATAAGCTCCGCTGAGCACCTCCAAGTTGTAGCCGTAAAGCTCTCTGTTTGTG 847
DB 842 TGTGAAATAATTAAGAGCGGTGTTGAATATCGACGCGTTAAAGCTCCAGGCTTTGGG 901
QY 848 AACGCGTAAAGCTATGCTTGAAGATATTCTATCTTACTCTGAGGAGAGCAATATTG 907
DB 902 ACAGAGAAAAAGAAATGCTCAAGACATCGCTATTTTAACCGCGCGTCAAGTCATTAGCG 961
QY 908 AAGATCGTGTATATAAGCTTGAAGATTAAGCTTCTTCTTTTGAAGACAGCTAAACGCTG 967
DB 962 AAGAATTTGGCTTGTAGTCTAGAAAAACGCTGAAGTGGAGTTTTTGAAGCAAGCTGGAAGA 1021
QY 968 TAGTTATTGACAAAGAAATACTACTATGTTGATGTTGCTGCAAAATCAGAAATCAGAAATTA 1027
DB 1022 TTGTGATTGACAAAGACCAACACCATGCTAGATGGCAAGGCCATAGCATGATGTTA 1081
QY 1028 AAGCTCGAGTTAAACAAATTCGTGCAAAATTTGAAGAAACAAAGCTCAGATTTATGATCGTG 1087
DB 1082 AAGACAGAGTTCGCGCAGATCAAAACCCAAATTTGCAAGTACGACAAAGCGATTTATGACAAA 1141
QY 1088 AAAAATTTCAAGAACTCTTGCAGAACTTGTGGTGGAGTACTGTTATCCATGTTGGAG 1147
DB 1142 AAAAATTTGCAAGAAATTTGGCTAACTCTCTGGCGGTGGCTGTGATTAAGTGGCG 1201
QY 1148 CTGCTACTGAAACTGAAATGAAGAGAGAGAGGATCGTGTAGAGATGCTCTAAATGCAA 1207
DB 1202 CTGCGAGTGAAGTGAATGAAG 1261
QY 1208 CAAGAGCTCGGTTGAAGAGGATTTGCTCCTGTTGGTGGTACTGCTTTTGTCCGCTCCA 1267
DB 1262 CTAAGCGCGGTTGAAGAGGATTTGATTTGGTGGCGTGGCGTCTCTCAITTCGCGCG 1321
QY 1268 TTAAGTCTTCTGATGATTTAAACCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1327
DB 1322 CTCAAAAAGTGCATTTGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1375
QY 1328 TCCGCTGCTTCTCTGAGAGCGCTTTACGTCATAATTTGCTCAAAATGCTGCTATGAGGTT 1387
DB 1376 TCATGCGGCCATTAAAGCCCATTTAGCTCAAAATCGCTATCAAGCTGCTTATGATGGCG 1435
QY 1388 CTATTGTTGAGAAAAAGTTGCTGAACCAAGAGAGGTTTGGATTTATGCTGATCAG 1447
DB 1436 GTGCTGCTGATGAATGAAGTAGAAAAACAGCAAGGCGCATTTTGGTTTTTAAACGCTAGCAATG 1495
QY 1448 GAGAATATGAAGACCTTATTAAAGCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1507
DB 1496 GCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1555
QY 1508 CATTACAAATCAGCATCAGTAGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1567
DB 1556 CTCACAAATGCGGTTTCGGTTTCAAGCTGCTTTTAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1615
QY 1568 AAAACCCAGAACTTAAAGAGATATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1624
DB 1616 AATCAAGAGAAAAAGCGACTCCGCAATGCTGATATGGTGGCATGGGCGGTATG 1675
QY 1625 GTGGTATGACGCTATGATTA 1646
DB 1676 GAGGATGGCGCATGATGTA 1697

RESULT 13

US-09-134-001C-1868

; Sequence 1868, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1868
LENGTH: 1623
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1868

Query Match 38.9%; Score 640; DB 4; Length 1623;
Best Local Similarity 63.1%; Pred. No. 9e-159;
Matches 1020; Conservative 0; Mismatches 590; Indels 6; Gaps 2;

QY 8 CTAAGAAATCCTTTTGTATCTAAAGCCCGTGAATAAATCTTACGAGGTGTAGATAAAC 67
DB 8 CAAGAATCTTAATCTCTGAAGATGCGGTCAAGCAATGTACGTGGTGTGATAAAT 67
QY 68 TTGCAATGCTGTTAAAGTAACACTTGGACCTAAAGCCGCTAAATGTCGTTAATGAAAAGT 127
DB 68 TAGCAACCGCTGTAAGGTTACAAATGGACCTAAAGCGGAAATGTTGTTCTAGATAAG 127
QY 128 CTTTGTGTTCCCACTGTTATTAACAAAGATGGTGATCTGTTGCAAAAGAAATGAACCTG 187
DB 128 ATTAACAACACACCTTTAATACCAACGATGTTGTAACAAATGCTAAAGAAATAGATTAG 187
QY 188 AAGATAAGTTTGAATAATGGCGCTCAAAATGTTAAAGAGTAGCTCCCAAAATAGCG 247
DB 188 AAGATCCATATGAGAAATGCGTCAAAATGTTAGCGAAGTTGCGAAATAAACAAATG 247
QY 248 ATATGCTGTGTGTAAGTAACACTACACAGCAACAGTCTTGCACAAAGCTATTTATCGTGAAG 307
DB 248 AAATCGCTGGGAGGTTACAACTACACAGCAACAGTCTTGCACAACTCAATGATTCAGGAAG 307
QY 308 GTGTAAACCTGTAGAGCTGTTGATCTATCTATGCGCAATTAACAGTGGCATAGATAAAG 367
DB 308 GTCTTAAGAAATGTACAAAGTGGTCAAAATCTGTAGGCTTAAAGCAAGTATTGACAAAG 367
QY 368 CTGTGTGTTGCTTACTAAAGACTAAGCGACATTAACAAAGCTCTCTGTCGACCAAAAG 427
DB 368 CAGTGAAGTGGCTATAGAAGCGCTTCATGAGATTTCTCAAAAGGTTGAAATTAAGAACG 427
QY 428 AATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACAACTAGGTAATATCATAG 487
DB 428 AGATAGCGCAAGTTGGAGCTATTTCAGCA---GCAGATGAAGAAATCGGTGCTACATTT 484
QY 488 CTGAAGCTATGCTTAAAGTTGAAAGAGAGGTGTTATCAAGTTGAGGAAGCTAAAGGTC 547
DB 488 CTGAAGCAATGTAAGTAGTAGTAACAGTGGCGTTATCACTATTGAAGATCAAAATCGGT 544
QY 548 TTCAAACTACATTAAGATGCTGTTGAAGGAATGAAGTTGACCGTGGCTACTCTCTCCAT 607
DB 548 TTAAATACAGATTAAGATGTTGAGGAAGTGAAGTTGATCGCGGTATCAATCAATCACCAT 604
QY 608 ACTTTTAACTAATCTCTGAGAAATGTTTGTGAACCTTGATAACCCCTTATATCTCTTGTGA 667
DB 608 ATATGTTAACTGACTCAGATAAATGATAGTCAATTAAGACGTCATATATATTAGTAA 664
QY 668 ATGAGAAAGATTAATAGCATGAAAGACATGCTCAAAATCTTAGAACAGTGTGCTAAAG 727
DB 668 CGGATAAGAAATTTTCAATCAATCCAGATATTTCTTCCATTTATTAAGAACAGTGTGAGG 724
QY 728 TAAACGCTCCACTCCTTATTTATGCTGAAGACGTAAGGTGAAGCACTTGAACACTTG 787
DB 728 CTAGTCAGCAATTTTAAATTTGTCGGATGAAGTAGAAGCGATGCACTTACTTAATATTG 784

QY 788 TAGTCAATAAGCTCGTGGAGCACTCCAAGTTGTTAGCCGTAAAGACTCTCTGTTTGGTG 847
DB 788 TTTTAAACCGTANGCGTGGACATTTACTGCTGTAGCAGTTTAAAGCCCGAGATTGGTG 844
QY 848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTTATCTTACTGAGGAGCAATATTG 907
DB 848 ATCGACGTAAGCAATGTTAGNAGCACTAGCAATATTAACTGGTCTCAAGTCATTACTG 904
QY 908 AAGATCGTGGTATAAAGCTTGAAGATGTAAGCTTCTTCTTTAGGAACAGCTAAACG 967
DB 908 ATGATTTAGGTTTAAAGCTTAAAGATGATCTCTTGTATGATGCTAGTACTGCTAATAAG 964
QY 968 TAGTTATTGACAAAGAAATATCTACTCTCTGTTGATGCTGCTGGAAGAAATCAGAGATATA 1027
DB 968 TTGAAGTCACTAAGATCATCAACAGCTGAGATGTAAGTGGTGAATGATAATAATATTG 1024
QY 1028 AAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGATCGTG 1087
DB 1028 ATGCTCGTGTAGTCAAAATTAAGACCAAAATTAAGAACTGATTCAGAGTTGATAAAG 1084
QY 1088 AAAAATCTCAAGAACGCTTTCGCAAACTTGTGTTGGAGTAGCTGTTTATCCATGTTGAG 1147
DB 1088 AAAAATTTACGAAAGTTTGGGAAACTTACCTGGGGGGTGTGTTTAAATCAAGTATCGG 1144
QY 1148 CTGCTACTGAACTGAATGAAGAGAGAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
DB 1148 GTGCACTGAACAGAGCTTAAAGAACGTAATTAAGAAATTAAGAGACGCAATTAATTCAA 1204
QY 1208 CAAGAGCTGGGTTGAAGAGAGTATTGCTCCTGCTGTTGGTGTGCTGTTTCTGCTGCCCTCA 1267
DB 1208 CAGCTGGCGGTTGAAGAGAGTATTGCTGCTGTTGGTGTGCTGTTTCTGCTGCCCTCA 1264
QY 1268 TTAAGTCTTGTATGATTAATTAACCTGCTGATGATGATGAATCTTGTGCTGCTTAAATCA 1327
DB 1268 ATCAAAAGTAAGTGAATTAAGAGCAGAGTGTATGTTGAAA---CGGGTGTAAATATCG 1321
QY 1328 TCCGCTGTTCTTGAAGAGCTTTAGCTCAAAATGCTGCAAAATGCTGCTGCTGCTGCTGCTG 1387
DB 1328 TATTAAGCATTTACAAGCACCTGTTTAGACAAATTTGCTGAAATGCAAGATTTAGAGGTT 1381
QY 1388 CTATGTTGTAGAAAAAGTTTCGTGAACCAAAAGATGTTTGGATTTAAATGCTGCATCAG 1447
DB 1388 CAATATTGTTGAACGTTTAAACATGCTGAAGCGCGGTTGGTTTCAATGACAGCAAA 1441
QY 1448 GAGATGAAGACCTTATTAAAGCTGGTGTCAATGATTCCTTAAAGAAAGTTTACAGTATTG 1507
DB 1448 ATGAATGGGTTAATATGTTAGAAGAGTATAGTAGATCCAACTAAAGTAACTGTTTCAAG 1501
QY 1508 CATTACAAATGAGCATCAGTACGCTCTTACTTCTAATCAGAAATGCTGCTGCTGCTGCTG 1567
DB 1508 CGTTACAACATGAGCAAGTGTAGTGTCTGTTTAACTGAACTGAAAGCACTGCTGCTGCTA 1561
QY 1568 AAAAACCAGAACCTTAAAGAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1623
DB 1568 GTATCCAGAGCCAGAAATTAAGAACCACTGGAATGGTGGATGCCAGGATG 1617

RESULT 14

US-09-221-017B-895/c
; Sequence 895, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 895:

SEQUENCE CHARACTERISTICS:

LENGTH: 2416 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORYPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc feature

LOCATION: 1...2416

US-09-221-017B-895

Query Match 38.0%; Score 626.4; DB 4; Length 2416;
Best Local Similarity 62.0%; Pred. No. 4.1e-155;
Matches 1009; Conservative 0; Mismatches 616; Indels 3; Gaps 1;

QY	8	CTAAAGAAATCCCTTTTGTATGCTAAAGCCCGTCAAAAACCTTTCCAGAGGTGTAGATAAAC	67
DB	1924	CAAAAGAAATCAATTCGATATGATCTCGACCTCTGAAGAGGGGTATAGTCAC	1865
QY	68	TTGCAAAATGCTGTAAAGTAACACTTGGACCTAAAGCCCGTAATGTGTTATTGAAAGT	127
DB	1864	TGCAAAATCGGTAAAGTTACCTCGGCGGAAAGGTGTAATGTTATCTTAGCAAGA	1805
QY	128	CTTTTGGTCCCGAGTTATTAACAAAGATGTTGTTCTGTCAAAGAAATTAAGACTTG	187
DB	1804	CGTACGAGCTCCGACATTAACCAAGACCGTGTGACCGTAGCGAAAGATAGATTTGG	1745
QY	188	AAGATAAGTTTGAATAATGGGCGCTCAAAATGTTAAAGAAAGTAGCTCCCAAAATAGCG	247
DB	1744	AGTGCCCGTTCGAAACATGGTGCCAGTTGGTGAAGAAAGTAGCTCCCAAGCAATG	1685
QY	248	ATATTGCTGTGTATGGAACATAACACAGCAAGTCCTTGGCAAGCTATTTATCGTGAAG	307
DB	1684	ACGATCCGCGTACCGTACGACTACGGCTACGATCTCGCCACAGCAATTTATCGGCGTG	1625
QY	308	GTGTAAACTGTAGAGCTGGTCTGTAATCTTATGGCCATTAACGTTGGCATAGATAAG	367
DB	1624	GTCTGAAGAACGTTACGCGAGGACCAATCGATGGGATTTGAACGGTGGTATCCACAGG	1565
QY	368	CTGTTGTTGCTTTACTTAAAGAACTAAGCGACATTACAAAGCCCTACTCGTACCAAA	424

DB	1564	CTGTAAGGCTGTGGTAACTCACAATTGCAGGTATGGCTAAGGAAGTGGCGACGACTTCC	1505
QY	425	AAGAATAAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAAACAATAGTAAATATCA	484
DB	1504	AGAATCGAGCAGTAGCCAAAGATCTCTGCCAACGGAGCAAAACATCGTAGCCTCA	1445
QY	485	TAGCTGAAGCTATGGCTAAAGTTGGAAGAGGAGTGTATACAGTTGAGGAAGCTAAAG	544
DB	1444	TTGCCGAGCCATGCGAAGGTGAAGAAAGAGCGTTATACGGTGAAGAAAGCAAGG	1385
QY	545	GTCTTGAACCTACATTAGATGTGGTTGAAGAAATGAAGTTTACCGTGGCTACCTCTCTC	604
DB	1384	GAAACGACACTACGGTAGAAGTGTGGAGGTATCGAGTTCCACCGCGCTACATCTCTC	1325
QY	605	CATACCTTTGTAATAATCTGAGAAATGTTTGTGAACCTTGATAACCCCTTATATCTTT	664
DB	1324	CCTACTTGTGAACGAAACCGATAGATGGAGTCAATGAAATCTTTTCACTCTCA	1265
QY	665	GTAATGAGAAAGAAATTAAGTACATGAAGACATGCTACCAATCTTTAGAAAGTTGCTA	724
DB	1264	TCTACGACAAAGAAATATCCGTCCTGAAAGAGATGCTCCGATCTCTGAAACAAACGGTTC	1205
QY	725	AAGTAAACCGTCACTCCTTATTTATGCTGAAGACGTGAAGTGAAGTGAAGTGAAGTGAAG	784
DB	1204	AGACGGGCAAGCCCTCTCTCATATTGCAAGACATCGACAGCGAAGCATCGCCACCT	1145
QY	785	TTGTAGTCAATAAGCTCCGTCGAGCACTCCAAGTTGTAGCGTAAAGCTCTCTGTTTGG	844
DB	1144	TGGTTGTAACCGCTCTGCGGCGAGCCTCAAGATCTGTGAGTGAAGGCTCCCGGATTCG	1085
QY	845	GTGAACGCGCTAAAGCTATGCTTGAAGATATTGCTATCTTACTTGAGAGAAAGCAATAT	904
DB	1084	CGCATCTGCAAGGCTATGCTGCAAGACATTTGCTATCTGACGGCGGAAACGGTTATCA	1025
QY	905	TTGAAGATCGTGTATTAAGCTTTGAAATGTAAGCTTGTCTCTTTTAGGAACAGCTAAAC	964
DB	1024	GCGAAGAGACCGCTCTGAAGCTCGAAATGCTACGATGGATATGCTCGGTACAGCTGAGA	965
QY	965	GTGTAGTATTGCAAGAAATACTACTATCTTGTATGCTGTGCTGGAATAATCAGAAGATA	1024
DB	964	AAGTAAACCGTTGATAAGGACATATCTATCTGTTAACGGAGCGGAAACAAAGAGGCA	905
QY	1025	TTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGAAGAAACAAAGCTCAGATTATGATC	1084
DB	904	TCGCTTACGATACACGAGATCAAAGCTCAGATCAGAAATACGACAGGACTACGACC	845
QY	1085	GTGAAACCTTCAAGAACGCTTGTGCAAACTTGTGTTGGTGGAGTACGTTATCCATGTTG	1144
DB	844	GCGAAAGAGCTGCAAGAACGTTTGGCCAAAGCTCCCGCGGTGTAGCTGTTCTTTACGTGG	785
QY	1145	GAGCTGCTACTGAAACTGAAATCAAGAGAAAGAGGATCGTGTAGAAGATGCTCTTAATG	1204
DB	784	GTGCTCCGAGAGTGGAAATGAAGAAAGAGGATCGCTGAGAAGATGCTTTCGATG	725
QY	1205	CAACAAGAGCTCGGTTGAAGAGGATTTGTCCTCGTGGTGGTACTGCTTTTGTTCGCT	1264
DB	724	CAACGGCTGCTCAATCGAAGAGGGTACAGTACCGGTGGCGGTACGCGCATACATTCGTG	665
QY	1265	CAATTAAGTCTTGTATGATATTAAACCTCTGATGATGATGATGATGATGATGATGATGAT	1324
DB	664	CCATAGTCTTTTGAAGGTTCTCAAGGGTGAGAACGAAAGACGAAACCCACAGGTATCGAGA	605
QY	1325	TCATCCGTCCTTCTTGTGAAGAGCTTTAGCTCAAAATTTGCTGCAAAATTTGCTGCTATGAAG	1384
DB	604	TCGTGAACCGCCATCGAGAGCGCTTCTGTCAGATCGTAGCGAACCGCGTAAAGAGG	545
QY	1385	GTCTTATTTGTAGAAAAAGTTGCTGAACCAAAAGATGTTTGGATTTTAACTGCTGAT	1444
DB	544	GTGCGGTTGTGTACAGAAGGTGAAGAGGAGGAGGAGCTTCGGCTTACATCCCGGTA	485
QY	1445	CAGGAGATATGAAGACCTTATTAAGCTGTGTCTTGTGATCTCTTAAAGAGTTTACACGTA	1504

Db 484 CGGATGTTTTCGAAAACCTCTACACTACCGGTGTTATCGACCGCGCAAGTAACACGCG 425
Qy 1505 TTGCATTAAATGAGCATCAGTACGCTCTTACTTCTTAACACTACAGAAATCGGCTATTG 1564
Db 424 TAGCATTGAAATGAGCATCAGGCTCTTACGAGGTATGTTCTGACTACGGAGTGGTTATCG 365
Qy 1565 CTGAAAACCGAAGCTTAAAGATATGCTGATGCTGCGGTGATGCGGTGATGCGGTGATGCG 1624
Db 364 CTGCAAGAAAGAGATATCTGCGGACCGGCAATGCGCGGAGGTATGCGGAGATGG 305
Qy 1625 GTGGTATG 1632
Db 304 CGGCAATG 297

RESULT 15
US-08-467-822-28
; Sequence 28, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2284 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-467-822-28

Query Match 37.1%; Score 610.4; DB 2; Length 2284;
Best Local Similarity 62.1%; Pred. No. 6.5e-151;
Matches 1019; Conservative 0; Mismatches 611; Indels 12; Gaps 3;

Qy 8 CTAAGAAATCCTTTTGTATGCTAAAGCCGCGTGAAGAACTTTTCAAGAGTGTAGATAAAC 67
Db 510 CAAAAGAAATCAAAATTTTTCAGATAGCGCAAGAAACCTTTTATTTGAAGCGGTAAAGCAAC 569
Qy 68 TTGCAAAATGCTGTTTAAAGTAACACTTGGACCTTAAAGGCGGTAAATGCTGTTTATGAAAGT 127
Db 570 TCCATGACGCTGTCAAAGTAACCATGGGCGCAAGAGGCGAGGAAACGTTGTTGATCCAAAAA 629
Qy 128 CTTTGTGTTTCCCGATTTATACAAAAGATGTTGTTATCTGTTGCAAAAGAAATTTGAACATTG 187
Db 630 GCTATGGCGCTCCAAGCATACCAAAGACGGGTGAGCGTGGCTTAAAGAGATTGAATTA 889
Qy 188 AAGATAAGTTTGAAGATATGGCGCTCAATGTTTAAAGAGTAGTCTCCCAAACTAGCG 247
Db 690 GTTGCCCGCTGGCTAAACATGGGCGCTCAGCTCGTTTAAAGAGATGCGAGCAAAACCGCTG 749
Qy 248 ATATTGCTGGTGTATGGAATACAAACAGCAACAGCTCTTGAACAGCTATTTATCGTGAAG 307
Db 750 ATGCGCGCGGATGGACGACGACGCGCTGGCTTATAGCAATTTTAAAGAGG 809
Qy 308 GTGTAAACTTTGTAGCAGCTGCTGTTATCTTATGCGCATTTAAAGCTGCGCATAGATAAG 367
Db 810 GCTTGAGGAATATCAGCGCTGGGCTTAACCTATTGAGTGAACGAGGCAATGATAAG 869
Qy 368 CTGTTGTTGTTTACTTAAGAACTAAGCGCATTTACAAAGCCCTACTCGTGACCAAAAAAG 427
Db 870 GCGCTGAAGCGATCATTAATGAGCTTAAAGAGCGAGCAAAAAAGTGGCGGTAAAGAG 929
Qy 428 AAATAGCTCAAGTTGGAACCACTTCTGCAACTCTGATACAACTAGTAAATATCATAG 487
Db 930 AAATCACCAGTAGGACCACTTCTGCAACTCCGATCACAAATATCGGGAATCATATCG 989
Qy 488 CTGAAGCTATGGCTTAAAGTTGGAAGAGGAGTGTATCACAGTTTGAGAGCTTAAAGTC 547
Db 990 CTGAGCTATGGAAGAAAGTGGTAAAGACGCGGTGATCACCGTTGAAGAGCTAAGGCA 1049
Qy 548 TTGAAGCTATGATGTTGTTGAAGAAATGAGTTTGAACGCTGCTCTCTCTCAT 607
Db 1050 TTGAAGATGAATTAGATGCTGTAAGGCGATGCAATTTGATAGAGGCTACCTCTCCCTT 1109
Qy 608 ACTTTGTAACCTAATCTCTGAGAAATGTTTGTGAACTTGTATACCTTATATCTTTGTA 667
Db 1110 ACTTTGTAACCAAGCTGAGAAATGAGCTGCAATTTGATAGAGCTTATATCTCTTTAA 1169
Qy 668 ATGAGAAAAAGATTACTTAGCATGAAGACATGCTTACCAATCTTGAACAAAGTTGCTAAG 727
Db 1170 CGGATAAAAAAATCTCTAGCATGAAGACATTTCCCGCTACTAGAAAAACCATGAAG 1229
Qy 728 TAAACCGTCCACTCTTATTTATGCTGAAGAGCTAGAGTGAAGCACTTTGCAACACTG 787
Db 1230 AGGCAAAACCGCTTTTAAATCATCGTGAAGACATTTGAGGCGGAGAGCTTTAACGACTCTAG 1289
Qy 788 TAGTCAATAAGCTCCGTTGAGCACTCCAAGTTCTAGCGTAAAGCTTCTGTTTGTGCTG 847
Db 1290 TGTGTAATAATTAAAGCGGTGTTGAATATCCAGCGGTAAAGCTTCCAGGCTTTGGGG 1349
Qy 848 AACCGCGTAAAGCTATGCTTTGAAGATATGCTTACTGAGGAGGAGCAATATTG 907
Db 1350 ACAGGAGAAAAAGAAATGCTCAAGACATCGCTGTTTAAACCGCGGTCAAGTCAITTAGCG 1409
Qy 908 AAGATCGTGTATAAGCTTGAAGATGTAAGTGTGCTTTCTTTAGGAACAGCTTAAACGCTG 967
Db 1410 AAGAAATGGGCTTGAGTCTAGAAAAACGCTGAAGTGGAGTTTTTTAGGCAAAAGGAA---GA 1466
Qy 968 TAGTTATTGACAAAGAAATATCTACTATCTGTTGATGCTGCTGGAATATCAGAGATATTA 1027
Db 1467 TTGTGATTGACAAAGACAAACCAAGATCGTAGTGGCAAGGCGCATAGCCATGAGCTCA 1526
Qy 1028 AAGCTCGAGTTAAACAAATTCGTGACAAATTTGAAGAAACAAAGCTCAGATTATGCTG 1087
Db 1527 AAGACAGAGTCGCGCAATCAAAACCCAAATTTGACAGCACGACGAGCGATTACGCAAAAG 1586

us-09-077-574a-1.rni

1088	AAAACTTTCAAGAACGTCCTTGCAAAACATGTTGGTGGAGTAGCTGTGTATCATGTTGGAG	1147
1587	AAAAATTCAGAAAAGATGGCCAAACTCTCTGGCGGTGTGCTGTGAATTAAGAGTGGCGG	1646
1148	CTGCTTACTGAACCTGAATGAAGAAGAGAGAGGATCGGTGAGAACATGCTCTAAATGCAA	1207
1647	CTGCGGAGTGAAGTGAATGAAGAAGAGAAAAGACGCGGTGGATGACCGCTTGACCGGGA	1706
1208	CAAGAGCTCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA	1267
1707	CTAAGCGCGGTTGAGAGAGCAATGTGATTTGGGGCGGTGCGGCCCTCATTCGCGCGG	1766
1268	TTAAGTCCCTTGATGATATTAACACCTGCTGATGATGAACCTTGGTGGACTTATATCA	1327
1767	CCCAAAAGATGCATTTGAATTTTAC-----ACGATGATGAAAAGGTGGGCTATGAANAATCA	1820
1328	TCCGTCGTCTCTTGAAGAGCCCTTTACGTCAAAATGCTGCAATGCTGGCTATGAAGGTT	1387
1821	TCATGCGCGCCATTAAAGCCGCCATTAGCTCAATTCGCTATCAATGCCGTTATGATGCGG	1880
1388	CTATTCTGTGAGAAAAAGTTTCGTGAACCAAAAAGATGTTTTCGATTTAATGCTGCATCAG	1447
1881	GTGTGCTCGTGAATGAAGTAGAAAAACACAGAGGGCATTTTGGTTTTTAACGCTAGCAATG	1940
1448	GAGAAATAGAGACCTTTTAAAGCTGGTGTCTATTCATCCTTAAAAAGTTACACGCTATG	1507
1941	GCAAGTATGGGACATGTTTTAAAGAAGGCATTATGACCCCTTTAAAGTAGAAGGATCG	2000
1508	CATTACAAAATGCAGCATCAGTACGCTCCTTACTTCTAACTACAGAAATCGGCTATTGCTG	1567
2001	CTTTTACAAAATGCGTTTCGGTTTTCAAGCTGCTTTTAAACCAGAGCCACCGTGCATG	2060
1568	AAAAACAGAACCTTAAAAAGATATGCTATGCGCTG---CGCGTGGTATGGGTGGTATGG	1624
2061	AAATCAAGAAGAAAAAGCGGCCCAAGCAATGCTGATATGGTGGCATGGGCGGAATGG	2120
1625	GTGGTATGAGCGGTATGACTA	1646
2121	GAGGCATGGGCGGCATGATGA	2142

Search completed: January 29, 2004, 15:27:33
Job time : 105 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:18:26 ; Search time 18.5 seconds
(without alignments)
8373.295 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 agggcttaagaatactct.....gtatggacgtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09077574/runat 28012004 125846 11606/app query.fasta_1.1799
-DB=SwissProt 41 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574 @CGN 1.1.18 @runat 28012004 125846 11606 -NCPV=6 -ICPV=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1891.5	63.7	547	1	CH60_RALSO
2	1890	63.3	546	1	CH60_ACSAC
3	1872.5	63.1	546	1	CH60_BURCE
4	1868.5	63.0	546	1	CH60_BURPS
5	1868.5	63.0	546	1	CH60_BURVI
6	1868	62.9	546	1	CH60_RHOPA
7	1867.5	62.9	546	1	CH60_BURTH
8	1867.5	62.9	547	1	CH60_BARHE
9	1867	62.9	545	1	CH61_RHIME
10	1866	62.8	551	1	CH60_AMOPS
11	1864.5	62.8	547	1	CH60_BARQU
12	1864	62.8	545	1	CH64_RHIME
13	1862.5	62.8	547	1	CH60_XYLEFA
14	1852.5	62.4	540	1	CH60_RHOMR
15	1852.5	62.4	546	1	CH60_CHRVI
16	1847	62.2	546	1	CH60_XANAVI
17	1845	62.2	545	1	CH60_RHOCA
18	1844.5	62.1	546	1	CH60_RHILV

19	1844	62.1	546	1	CH60_XANCH
20	1843.5	62.1	547	1	CH60_BORPE
21	1843.5	62.1	552	1	CH60_COXBU
22	1842	62.1	546	1	CH60_XANCP
23	1841	62.0	542	1	CH65_RHIME
24	1840.5	62.0	544	1	CH60_AGRIT5
25	1834.5	61.8	546	1	CH61_RHOSH
26	1834.5	61.8	552	1	CH60_PSEST
27	1834	61.8	542	1	CH62_RHILE
28	1833.5	61.8	539	1	CH60_PARDE
29	1832.5	61.7	545	1	CH60_BRAJA
30	1828.5	61.6	545	1	CH60_CHLTE
31	1826	61.5	542	1	CH62_RHILO
32	1825.5	61.5	547	1	CH60_PSEAE
33	1823	61.4	546	1	CH60_BRUBA
34	1819	61.3	546	1	CH60_BRUSU
35	1818	61.3	549	1	CH62_BRAJA
36	1816	61.2	542	1	CH62_RHIME
37	1814.5	61.1	544	1	CH60_NEIMA
38	1814	61.1	546	1	CH60_BRUME
39	1812.5	61.1	551	1	CH64_RHILO
40	1810	61.0	552	1	CH63_RHILO
41	1808.5	60.9	544	1	CH60_NEIMB
42	1805.5	60.8	543	1	CH60_BARBA
43	1804.5	60.8	547	1	CH60_ALTHA
44	1803.5	60.8	549	1	CH65_RHILO
45	1802.5	60.7	544	1	CH60_NEIGO

ALIGNMENTS

RESULT 1

ID	CH60_RALSO	STANDARD;	PRT;	547 AA.
AC	Q8Y1P8;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	60 kDa chaperonin (Protein Cpn60) (groEL protein).			
GN	GROEL OR GROEL OR MOPA OR RSC0642 OR RSO1546.			
OS	Ralstonia solanacearum (Pseudomonas solanacearum).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC	Burkholderiaceae; Ralstonia.			
OX	NCBI_TaxID=305;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GM11000;			
PX	MDL=21681879; PubMed=11823852;			
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangelot S.,			
RA	Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,			
RA	Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,			
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,			
RA	Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,			
RA	Weissenbach J., Boucher C.A.;			
RT	"Genome sequence of the plant pathogen Ralstonia solanacearum";			
RL	Nature 415:497-502(2002)			
CC	-!- FUNCTION: Prevents misfolding and promotes the refolding and			
CC	proper assembly of unfolded polypeptides generated under stress			
CC	conditions (By similarity).			
CC	-!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of			
CC	7 subunits (By similarity).			
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-!- SIMILARITY: Belongs to the chaperonin (HSP60) family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; AL646060; CAD14172.1; --			

DR HAMAP; MF 00600; -, 1.
 DR InterPro; IPR001844; Chaprinin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR Pfam; PF00118; Cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONIN60; 1.
 DR Chapterone; ATP-binding; Complete proteome.
 SQ SEQUENCE 547 AA; 57404 MW; 0487E4650867C25A CRC64;

Alignment Scores:

Pred. No.: 2,328-97 Length: 547
 Score: 1891.50 Matches: 380
 Percent Similarity: 82.27% Conservative: 70
 Best Local Similarity: 69.47% Mismatches: 96
 Query Match: 63.73% Indels: 1
 Ds: 1 Gaps: 1

US-09-077-574A-1 (1-1647) x CH60_RALSO (1-547)

QY 1 ATGGCTTCTAAAGAAATCTTTTTCATGCTAAAGCCGCTGAAAGAACTTTTCACGAGGTGA 60
 Db 1 MetAlaAlaLysAspValPheGlyAspAlaAlaArgAlaLysMetValGluGlyVal 20
 QY 61 GATAAATCTGCAATGCTGTTAAAGTAACACTTGGACCTTAAAGCGCGTAATCTCGTTAT 120
 Db 21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
 QY 121 GAAAGCTCTTTGGTTCCCGAGTATTACAAAGATGGTGATCTGTTGCAAGAAAT 180
 Db 41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60
 QY 181 GAACCTTGAGATAAGCTTTGAAATATGGCGCTCAAAATGGTTAAAGAGTAGCTCCCAA 240
 Db 61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
 QY 241 ACTAGCATATTTGCTGGTGAAGTCAACAGCAACAGTCTTTCACAGCTATTTAT 300
 Db 81 ThrSerAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
 QY 301 CGTGAAGTGTAAACCTTGACGCTGCTGCTATCTATGCGCATTAACAGTGGCATA 360
 Db 101 ArgGluGlyMetLysTyrValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
 QY 361 GATAAAGCTGTGTGCTGTTACTTAAAGAACTTAAAGCGACATTAACAGCTACTCGTGC 420
 Db 121 AspLysAlaValAlaAlaValGluGluLeuLysLysIleSerLysProThrThrThr 140
 QY 421 CAAAAGAAATAGCTCAAGTTGGAAACCACTTCTGCAAACTCTGTACACAACTAGGTAAT 480
 Db 141 SerLysGluIleAlaGlnValGlyAlaIleSerAlaAsnSerAspGluSerIleGlyAla 160
 QY 481 ATCATAGCTGAGCTATGCTTAAAGTGTGAAGAGAGGTGTTATCACAGTTGAGGAAGCT 540
 Db 161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
 QY 541 AAAGCTCTTGAACCTACATCTAGCTGTGTTGAAGAACTCAAGTTTGACCGTGGCTACCTC 600
 Db 181 LysSerLeuGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
 QY 601 TCTCATCTTTGTAACTAATCTCAGAAATGGTTTGTGAACTTTGATAACCTTATATC 660
 Db 201 SerProTyrPheIleAsnAsnProGluLysGlnValValGlnLeuAspAsnProPheVal 220
 QY 661 CTTTGTATGAGAAAAGATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 720
 Db 221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
 QY 721 GCTAAAGTAAACCGCTCCCTCTTATTGCTGAAGAGCTGAGAAGGTGAAGCACTTGA 780
 Db 241 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260
 QY 781 ACACCTTAGTCAATAAGCTCGTGGAGCACTCCAAAGTTGTAGCCGCTAAAGCTCCTGCT 840

Db 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrAlaAlaValLysAlaProGly 280
 QY 841 TTTGGTGAACGCCGTAAGCTATGCTTGAAGTATATGCTTCTTCTGAGGAGAGCA 900
 Db 281 PheGlyAspArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
 QY 901 ATATTGGAAGATCGTGTATAAAGCTTGAAGATGTAAGCTTCTTCTTTAGCAACAGCT 960
 Db 301 IleAlaGluGluValGlyLeuThrLeuGluLysAlaThrLeuAsnAspLeuGlyGlnAla 320
 QY 961 AAACCTGTAGTTATTGACAAAGAAATATCTACTATCTGTTGATGGTCTGCTGAAATACAA 1020
 Db 321 LysArgValGluIleGlyLysGluAsnThrThrIleAspGlyAlaGlyAspAlaArg 340
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTAT 1080
 Db 341 AsnIleGluAlaArgValLysGlnValArgAlaGlnIleGluGluAlaThrSerAspTyr 360
 QY 1081 GATCTGAAAAAATCTTCAAGAACCTCTTGCAAAACCTTGTGGTGGAGTAGCTGTATCCAT 1140
 Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyIleValAlaValIleLys 380
 QY 1141 GTTGAAGCTGCTACTGAAACTGAAATGAAAGAAAGAGATCGTGTAGAGATGCTTA 1200
 Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
 QY 1201 AATGCAACAGAGCTGCGGTGCAAGAGTATTGTCCTGCTGGTGGTACTCTCTTTGTC 1260
 Db 401 HisAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyValAlaLeuLeu 420
 QY 1261 CGCTCCATTAAGTCTTGTATTAATTAACCTCTGATGATGATGATGATGATGATGATGAT 1320
 Db 421 ArgAlaArgAlaLeuIleSerGlyLeuLysGlyAlaAsnAlaAspGlnAspAlaGlyIle 440
 QY 1321 AATATCATCCGCTGCTTCTTGAAGAGCTTTTACGTCATAATGCTGCAAAATGCTGCTAT 1380
 Db 441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnAlaGlyAsp 460
 QY 1381 GAAGTTCTTATTTGTAGAAAAGTTCGTGACCAACAAAGATGGTTTGGATTTAATGCT 1440
 Db 461 GluAlaSerValValAlaAsnValIleAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480
 QY 1441 GCATCAGCAGAAATATGAAGACCTTATAAGCTGTGTCATTGATGATGATGATGATGATGAT 1500
 Db 481 SerThrGlyGlyTyrGlyAspLeuValGluMetGlyValLeuAspProThrLysValThr 500
 QY 1501 CGTATTGATTACAAAATGCAAGATCAGTACGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
 Db 501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrAspCysAla 520
 QY 1561 ATTGCTGAAAACCAACCAACCTTAAAAAGATATGCTTATCCCTGCGGTGGTATGGTGGT 1620
 Db 521 ValAlaGluLeuProLysAspAlaAlaProAlaMetPro---GlyGlyMetGlyGly 539
 QY 1621 ATGGTGTGCTATGACGCTATG 1641
 Db 540 MetGlyGlyMetAspGlyMet 546
 RESULT 2
 CH60_ACEAC
 ID CH60_ACEAC STANDARD; PRT; 546 AA.
 AC Q8GBD2;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
 GN GROEL OR GROEL.
 OS Acetobacter acetii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Acetobacteraceae; Acetobacter.
 OX NCBI_TaxID=435;
 RN [1]

QY 1618 GGTATGGTGGTATGGAC 1635
 |||||
 Db 540 GlyMetGlyGlyMetAsp 545

RESULT 3

CH60_BURCE STANDARD; PRT; 546 AA.
 AC Q92PE0;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60 kDa chaparonin (Protein Cpn60) (groEL protein).
 GN GROEL OR GROEL OR MOPA.
 OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia.
 OX NCBI_TaxID=292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NTC 10744;
 RA Zysk G., Spletstoesser W.D., Neubauer H.;
 RT "Nucleotide sequence comparison of the groE operon of Burkholderia
 spp.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
 proper assembly of unfolded polypeptides generated under stress
 conditions (By similarity).
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 7 subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaparonin (HSP60) family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

EMBL; AF104907; AAC79087.1; -.
 DR HSP; P06139; 1GRL.
 DR HAMAP; MF 00600; 1.
 DR InterPro; IPR001844; Chaprinin_Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR Pfam; PF00118; cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 KW Chapterone; ATP-binding.
 SQ SEQUENCE 546 AA; 56980 MW; 174B9934345E7315 CRC64;

Alignment Scores:
 Pred. No.: 2,59e-96 Length: 546
 Score: 1872.50 Matches: 374
 Percent Similarity: 82.08% Conservative: 75
 Best Local Similarity: 68.37% Mismatches: 95
 Query Match: 63.09% Indels: 3
 DB: 1 Gaps: 2

US-09-077-574A-1 (1-1647) x CH60_BURCE (1-546)

QY 1 ATGGCTCTAAAGAAATCCCTTTTGATGCTAAAGCCCGTGAAGAACTTTCACGAGGTCTA 60
 |||||
 Db 1 MetAlaAlaLysAspValPheGlyAspSerLysMetValGluGlyVal 20
 |||||
 QY 61 GATAAACTCGAATGCTGTTAAAGTAACACTTGGACCTAAAGCCCGTAATGCTGTTATT 120
 |||||
 Db 21 AsnLeuLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
 |||||
 QY 121 GAAAGTCTTTTGGTTCCCGCATTTATTAACAAAGATGGTGTATCTGTTGTCGAAAGAAATT 180
 |||||

Db 41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60
 QY 181 GAACCTTGAAGATAACTTTGAAATATGGCGCTCAAAATGCTTAAAGAACTAGCTCCCAA 240
 |||||
 Db 61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
 |||||
 QY 241 ACTAGCGATATTGCTGGTGAAGAACTACAACAGCAACAGCTCCTTGCACAACTATTAT 300
 |||||
 Db 81 ThrSerAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
 |||||
 QY 301 CGTGAAGGTGTAACACTTGTAGCAGCTGCTGTAATCCTATGCCATTAAGCTGCGATA 360
 |||||
 Db 101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120
 |||||
 QY 361 GATAAAGCTGTTGCTGTTACTAAAGAACTAAGCGACATTAACAAAGCTACTCGTAC 420
 |||||
 Db 121 AspLysAlaValAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140
 |||||
 QY 421 CAAAAGAAATAGCTCAAGTGGAAACCATTTCTGAACTCTGAACTCTGATACAACATAGTA 480
 |||||
 Db 141 AsnLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspSerSerIleGlyAsp 160
 |||||
 QY 481 ATCATAGCTGAAGCTATGCTAAAGTTGGAAGAGGAGGTGTTATACAGTTGAGGAAGCT 540
 |||||
 Db 161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
 |||||
 QY 541 AAAGGTCTTGAAGAACTACATAGATGTTGGAAGAAATGAAGTTTACCGTGGCTACCTC 600
 |||||
 Db 181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
 |||||
 QY 601 TCTCCATACTTTGTAATAATCCTCGAGAAAATGGTTTGTGAAGTCTGATTAACCTTATATC 660
 |||||
 Db 201 SerProTyrPheIleAsnAsnProAspLysGlnValAlaValLeuAspAsnProPheVal 220
 |||||
 QY 661 CTTTGTATGAGAAAGAAATTAAGTACATGAAAGACATGCTACCAATCTTAGAACAAAGTT 720
 |||||
 Db 221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
 |||||
 QY 721 GCTAAAGTAAACCGTCCACTCCTCTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGA 780
 |||||
 Db 241 AlaLysAlaGlyArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
 |||||
 QY 781 ACATCTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTCGT 840
 |||||
 Db 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280
 |||||
 QY 841 TTTGCTGAACCGCTAAAGCTATGCTGAAGATATTCCTATCCTTACTGAGGAGAGCA 900
 |||||
 Db 281 PheGlyAspArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
 |||||
 QY 901 ATATTTGAAGATCGTGTATATAAGCTTGAAGTGTAAAGTGTGTCTTCTTTAGGAACAGCT 960
 |||||
 Db 301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320
 |||||
 QY 961 AAACGTGTAGTTATTCACAAAGAAATATCTATCTGTTGATGCTGCTGGAATACAGAA 1020
 |||||
 Db 321 LysArgIleGluValGlyLysGluAsnThrThrIleIleAspGlyAlaGlyGluAlaAla 340
 |||||
 QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGACAAACAAAGCTCAGATTAT 1080
 |||||
 Db 341 SerIleGluAlaArgValLysGlnValArgAlaGlnIleGluGluAlaThrSerAspTyr 360
 |||||
 QY 1081 GATCGTGAAACCTTCAAGAACCTCTTGCACAACTTGTGTGGTAGGAGTAGCTGTATCCAT 1140
 |||||
 Db 361 AspArgGluLysLeuGlnValArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
 |||||
 QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
 |||||
 Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
 |||||
 QY 1201 AATGCCACAGAGCTCGGTTGAAGAGGTATTGTCCTCGTGGTGGTACTGCTGCTTTGTC 1260
 |||||
 Db 401 HisAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIle 420
 |||||

```
QY 1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGAACACTTGGACTT 1320
Db 1261 ArgAlaArgThrAlaAlaGlyLeuThrGlyAlaAsnAlaAspGlnAsnAlaGlyLe 440
QY 1321 AATATCATCGTGGTCTCTGGAAGAGCTTACGCTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db 1321 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGly 460
QY 1381 GAAGTCTCTATTGTTGTAAGAAAGCTTGTGAACCAAGAGCTGTTTGGATTATGCT 1440
Db 1381 GluAlaSerValValAlaAlaValAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480
QY 1441 GCATCAGGAGATATGAAGACTTATTAAAGCTGGTGCATTCATGCTCACTCAAAAGCTTACA 1500
Db 1441 AlaThrGlyGlyTyrValAspMetValGluAlaGlyValValAspProThrLysValThr 500
QY 1501 CGTATTGATTCACAAATGCGAGCTACGATGCTGCTCTTCTTAACTACAGAAATGCGCT 1560
Db 1501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520
QY 1561 ATTGCTGAAACACCAAGCTTAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1561 ValAlaGluLeu-----ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538
QY 1621 ATGGGTGGTATGACGGTATG 1641
Db 1621 MetGlyGlyMet---GlyMet 544

RESULT 4
CH60_BURPS STANDARD; PRT; 546 AA.
AC Q9F712;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR MOGA.
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=28450;
RN [1]
RP SEQUENCE FROM N.A.
RA Woo P.C.Y., Leung P.K.L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF287633; AAC32927.1; --
CC HSPSP; P06139; IGLR.
CC HMAP; MF_00600; --; 1.
CC InterPro; IPR001844; Chaperonin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1. 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC
CC -----
CC KW Chaperone; ATP-binding.
CC
CC -----
CC SQ SEQUENCE 546 AA; 57145 MW; E9E9366EE8BD6ABD CRC64;
```

```
Alignment Scores:
Pred. No.: 4,3e-96 Length: 546
Score: 1868.50 Matches: 373
Percent Similarity: 81.90% Conservative: 75
Best Local Similarity: 68.19% Mismatches: 96
Query Match: 62.95% Indels: 3
DB: 1 Gaps: 2

US-09-077-574A-1 (1-1647) x CH60_BURPS (1-546)
QY 1 ATGGCTCTTAAGAAATCTTTTTCATCTAAAGCCCTGAAATACTTTACAGAGGTGA 60
Db 1 MetAlaAlaLysAspValValPheClyAspSerAlaAlaLysMetValGluGlyVal 20
QY 61 GATAACTTGCCTCAATGCTGTTAAAGTAACTTGCAGCTAAAGCCGTAATTCGTATT 120
Db 21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValLeu 40
QY 121 GAAAGTCTTTTGGTCTCCAGTATTACAAAGATGCTGATCTGTGCAAAAGAAATT 180
Db 41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluLe 60
QY 181 GAACTTGAAGATAAGTTTGAATAATATGGCGCTCAAATGTTTAAAGAGTAGCTCCCAA 240
Db 61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATGTCTGTGATGGAACACTCAACAGCAACAGTCCTTGCACAGCTATTAT 300
Db 81 ThrSerAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100
QY 301 CGTGAAGTCTTAAACTTGTAGCAGCTGCTCGTAACTCTATGCGCATTAACCGTGCATA 360
Db 101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyLe 120
QY 361 GATAAAGCTGTGTGCTGTCTTACTTAAAGAACTTAAGCGACATTAACAAAGCCTACTCGTGAC 420
Db 121 AspLysAlaValAlaAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140
QY 421 CAAAGAAATAGCTCAAGTTGGAACCACTTCTCAAACTCTGATACAAACATAGGTAAT 480
Db 141 AsnLysGluIleAlaGlnValGlyAlaIleSerAlaAsnSerAspSerSerIleGlyAsp 160
QY 481 ATCATAGCTGAAGCTATGCTTAAAGTTGGAAGAGAGGTGTTATCACAGTTGAGAGAGCT 540
Db 161 ArgIleAlaGluAlaMetAspLysValGlyLysGlyValIleThrValGluAspGly 180
QY 541 AAAGGTCTTCAAACTACATTAGATGCTTGAAGGAATGAAGTTGACCGTGGCTACTCTC 600
Db 181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY 601 TCTCCATCTTTGTAATACTCTCTGAGAAATGTTTGTGAACCTTGATAACCCCTTATATC 660
Db 201 SerProTyrPheIleAsnAsnProAspLysGlnValAlaValLeuGluAsnProPheVal 220
QY 661 CTTTGTATCAGAAAAGATCTAGCATCAAAAGACATGCTACCAATCTTAGAACAGTT 720
Db 221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTTATTATTGCTGAAGACGTAGAAGGTGAAGCATTGCA 780
Db 241 AlaLysAlaGlyArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY 781 ACATTGTAGTCAATAAGCTCCGTGGAGCACTCCAAAGTTGTAGCCGTAAAGCTCTCGT 840
Db 261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGCTTAAAGCTATGCTTGAAGATATTGCTATCTTACTGGAGAGAGCA 900
Db 281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
QY 901 ATATTGTAAGATCGTGTATTAAGCTTGAAGATGTAAGCTTCTTCTTCTTAGAAACAGCT 960
```

```
Db 301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320
Qy 961 AAACGTGTAGTATTGCAAGAAATACTACTATCTGTTGATGCTGCTGCGAAATACAGAA 1020
Db 321 LysArgIleGluValGlyLysGluAsnThrThrIleAspGlyAlaGlyGluAlaVal 340
Qy 1021 GATATTAAGCTCGAGTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATTAT 1080
Db 341 AsnIleGluAlaArgValLysGlnIleArgThrGlnIleGluGluAlaThrSerAspTyr 360
Qy 1081 GATCGTGAATAACTTCAAGAACTCTTCAAACTTGTGGTGGAGTACTCTTATCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
Qy 1141 GTTGAGCTGCTACTGAACTGAATGAAGAGAGAGAGATCGTGTAGAGATGCTCTA 1200
Db 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
Qy 1201 AATGCAACAAGAGCTGCGGTTCAAGAGATTTGTCCTCTGCTGGTGGTACTGCTTTGTC 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuIle 420
Qy 1261 CGCTCCATTAAGTCTTCATGATATTAACCTGCTGATGATGATGATGATGATGATGAT 1320
Db 421 ArgAlaArgThrAlaIleAlaSerLeuThrGlyValAlaAlaAspGlnAsnAlaGlyIle 440
Qy 1321 AATATCATCGTCTGCTCTTGAAGAGCTTTACGTCAATTCCTCAATTCCTCAATTCCTCA 1380
Db 441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGlyGlu 460
Qy 1381 GAAGGTTCTATTGTTCTGACAAAGAGTTCGTGAACCAAGAGAGTTCGTTTGGATTATGCT 1440
Db 461 GluAlaSerValValAlaAlaAlaValAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480
Qy 1441 GCATCAGGAGATATGAGACCTTATTAAAGCTGGTGTCAATGATCCTAAAGATTACA 1500
Db 481 AlaThrGlyGlyThrValAspMetValGluAlaGlyValValAspProThrLysValThr 500
Qy 1501 CGTATTGCTTACAAATCAGATCAGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
Db 501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520
Qy 1561 ATTGCTGAAACACCAAGACCTTAAAGAGATATGCTCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 521 ValAlaGluLeu-----ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538
Qy 1621 ATGGGTGGTATGACCGGTATG 1641
Db 539 MetGlyGlyMet---GlyMet 544

RESULT 5
CH60_BURVI STANDARD; PRT; 546 AA.
AC Q9ZFD8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL OR MOPA.
OS Burkholderia vietnamiensis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=60552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 11319;
RA Zysk G., Splettoesser W.D., Neubauer H.;
RT "Nucleotide sequence comparison of the groE operon of Burkholderia
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC
```

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL; AF104908; AAC79089.1; -;
CC HSSP; P06139; 1GRU.
CC HAMAP; MF 00600; -; 1.
CC InterPro; IPR001844; Chaperin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 546 AA; 57010 MW; 47FA893959221218 CRC64;

Alignment Scores:
Pred. No.: 4,3e-96 Length: 546
Score: 1868.50 Matches: 373
Percent Similarity: 81.90% Conservative: 75
Best Local Similarity: 68.19% Mismatches: 96
Query Match: 62.95% Indels: 3
DB: 1 Gaps: 2

US-09-077-574A-1 (1-1647) x CH60_BURVI (1-546)

Qy 1 ATGGCTTCTAAGAAATCCTTTTGTGATGCTAAAGCCGCTGAAACATTTTCACGAGGTGA 60
Db 1 MetAlaAlaLysAspValValPheGlyAspSerAlaArgSerLysMetValGluGlyVal 20
Qy 61 GATAAATCTGCAATCTGTTAAAGTAACACTTGGACCTTAAAGCCGCTAATGTCGTTATT 120
Db 21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
Qy 121 GAAAGCTCTTTTGGTTCCCGCTTATTACAAAAGATGGTGTATCTGTGTCGAAAGAAATT 180
Db 41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60
Qy 181 GAACCTTGAGATTAAGTTTGAATATGCGGCTCAATGGTTAAAGAGTAGTCCCAAA 240
Db 61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaLysLys 80
Qy 241 ACTAGCGATATTCTGCTGATGCACTTACAAACAGCAACAGCTCTCTGCAAGCTATTAT 300
Db 81 ThrSerAspAsnAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
Qy 301 CGTGAAGGTGTAATACTTGTAGCAGCTGGTGTGTAATCTCTATGCGCATTAACGTCGATA 360
Db 101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120
Qy 361 GATAAAGCTGTTGTTGCTGTACTTAAGAAGTAAGCGACATTACAAAGCCTACTCTGTGAC 420
Db 121 AspLysAlaValAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140
Qy 421 CAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAAACATAGGTAAAT 480
Db 141 AsnLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspSerSerIleGlyAsp 160
Qy 481 ATCATAGCTGAAGCTATGCTGCTTAAAGTTGGAAGAGAGGTGTTATCATAGTTGAGAGACT 540
Db 161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
Qy 541 AAAGCTTCTGAAACTACTATTAGATGTTGTTGCAAGGAATGAAGTTTGACCGGTGCTACCTC 600

```
Db 181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
QY TCTCATACCTTTGTAACCTAGAAATGTTTGTGAACCTGATGAACTTATATC 660
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
201 SerProTyrPheIleAsnAsnProAspGlyGlnValAlaValLeuAspAsnProPheVal 220
QY CTTTGTATGAGAAAGATTACTAGCATGAAGACATGCTACCACTCTTAGAACAACTT 720
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240
QY GCTAAAGTAAACCGTCCACTCTTATATTTCTGTAAGACGTAGAGGTGAAGCACTTGC 780
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
241 AlaLysAlaGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260
QY ACACCTGTAGTCAATAAGCTCCGTGGAGCACTCCAAAGTTGTAGCGTAAAGCTCCCTGGT 840
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280
QY TTTGGTGAACCGCTAAAGCTATCTTGAAGATATTGCTATCTTACTGGAGGAGAAGCA 900
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
281 PheGlyAspArgGlyAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
QY ATATTTGAAGATCGTGTATTAAGCTTGAATGTAAGCTGTCTTCTTTAGGAACAGCT 960
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320
QY AAACGTGTAGTATTGTAAGAAAGAAATACTACTATCTGTGTATGCTGGTGAAGATCAGAA 1020
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
321 LysArgIleGluValGlyLysGluAsnThrThrIleIleAspGlyAlaGlyGluAlaAla 340
QY GATATTAAAGCTGAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAAAGCTCAGATTAT 1080
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
341 SerIleGluAlaArgValLysGlnValArgThrGlnIleGluGluAlaThrSerAspTyr 360
QY GATCGTGAAGAACTCAAGACGCTCTGCAAACTTGTGTGAGTAGCTGTTATCCAT 1140
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY GTTGGAGCTGCTACTGAACTGAATCAAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
381 ValGlyAlaAlaThrGluValGluMetCysGluLysLysAlaArgValGluAspAlaLeu 400
QY AATGCAACAAGAGCTGCGGTGTAAGAGGATTTGTCCTCGTGTGTGCTACTGTTTGTGTC 1260
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
401 HisAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyValAlaLeuIle 420
QY CGTCCATTAAAGCTTGTATGATATTAACCTGCTGATGATGATGAACTTGTGGACTT 1320
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
421 ArgAlaArgThrAlaIleAlaGlyLeuThrGlyAlaAsnAlaAspGlnAsnAlaGlyIle 440
QY AATATCATCGTCTGTTCTCTTGAAGACCTTTAGTCAAAATTCGTCAAAATGCTGCTAT 1380
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGlyGlu 460
QY GAAGGTCTTATGTTGTAGAAAAGATTCGTGAACCAAGATGTTTGTGATTAATGCT 1440
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
461 GluAlaSerValValAlaAlaValAlaAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480
QY GCATCAGGAATATGAAGACCTTATTAAGCTGTTGTCATGATCTCAAAAAGTTTACA 1500
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
481 AlaThrGlyGlyTyrValAspMetValGluAlaGlyValValAspProThrLysValThr 500
QY CGTATTGATTACAAATGACGATCAGTACCTCTTACTTCTTAAGTACAGAACTCGCT 1560
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520
QY ATTGCTGAAACCAAGAACCTTAAAGATATATGCTATGCTGCGGTGCTATGCTGCTG 1620
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
521 ValAlaGluLeu-----ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538
QY ATGGGTGGTATGACGGTATG 1641
Db |||||...|||...|||...|||...|||...|||...|||...|||...|||...
539 MetGlyGlyMet---GlyMet 544
```

```
RESULT 6
CH60_RHOPA STANDARD; PRT; 546 AA.
AC Q93MH1
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROU OR GROEL.
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao M., Zhu C.R., Qian X.M., Zheng P., Chen Y.Y.;
RT "Cloning and Sequencing of the groEL Operon of Rhodopseudomonas
palustris.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF406639; AAK94943.1; -.
CC HAMAP; MF_00600; -.
CC InterPro; IPR001844; Chaprinin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONIN_CPN60; 1.
CC Chaperone; ATP-binding.
CC
CC SQ SEQUENCE 546 AA; 57937 MW; 163C7115D68FFFP8 CRC64;
CC
CC Alignment Scores:
CC Pred. No.: 4,59e-96 Length: 546
CC Score: 1868.00 Matches: 374
CC Percent Similarity: 81.93% Conservatve: 75
CC Best Local Similarity: 68.25% Mismatches: 95
CC Query Match: 62.94% Indels: 4
CC DB: 1 Gaps: 2
CC
CC US-09-077-574A-1 (1-1647) x CH60_RHOPA (1-546)
CC
CC QY 1 ATGGCTTCTAAAGAAATCTCTTTTGTATGCTTAAAGCCGCTGAAAACTTTCCAGGAGTGA 60
CC Db 1 MetAlaAlaLysAspValLysPheAspThrAspAlaArgAspArgMetLeuArgGlyVal 20
CC
CC QY 61 GATAAATCTGCAATCTGTTAAAGTAACACTTGGACCTAAAGCCGCTAATGCTGTTATT 120
CC Db 21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40
CC
CC QY 121 GAAAGCTCTTTGTTCCCGAGTATTATCAAGAGATGCTGCTATCTCTTCCAAAGAAATT 180
CC Db 41 AspLysSerPheGlyAlaProGlyIleThrLysAspGlyValSerValAlaLysGluIle 60
CC
CC QY 181 GAACCTTGAAGTAAGTTTCGAAATATGCGCGCTCAAAATGCTTAAAGAGTAGCTCCCAA 240
CC Db 61 GluLeuSerAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80
```


RT analysis tool.;

BL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-408 FROM N.A.

RC STRAIN=ATCC 49882 / Houston 1;

RX MEDLINE=97373904; PubMed=9230387;

RA Sumner J.W., Nicholson W.L., Massung R.F.;

RT "PCR amplification and comparison of nucleotide sequences from the

RT groSL heat shock operon of *Escherichia coli*,"

RL J. Clin. Microbiol. 35:2087-2092(1997).

CC -!- FUNCTION: Prevents misfolding and promotes the refolding and

CC proper assembly of unfolded polypeptides generated under stress

CC conditions (By similarity).

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

CC 7 subunits (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U78514; AAB69094.1; -; -; ALT_INIT.

DR EMBL; AF014829; AAD04238.1; -; -; ALT_INIT.

DR EMBL; U96734; AAB65637.1; -; -; -;

DR HSSP; P06139; 1GRL.

DR HAMAP; MF 00600; -; 1.

DR InterPro; IPR001844; Chaperonin Cpn60.

DR InterPro; IPR002423; Cpn60/TCP-1.

DR Pfam; PF00118; cpn60_TCP1; 1.

DR PRINTS; PRO0298; CHAPERONING60.

DR PRINTS; PRO0304; TCOMPLEXTCP1.

DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.

DR KQ Chaperone; ATP-binding; Heat shock.

SQ SEQUENCE 547 AA; 57625 MW; EDE25D566D5B10D2 CRC64;

Alignment Scores:

Pred. No.: 4,89e-96 Length: 547

Score: 1867.50 Matches: 385

Percent Similarity: 81.24% Conservative: 61

Best Local Similarity: 70.13% Mismatches: 96

Query Match: 62.92% Indels: 7

DB: 1 Gaps: 4

US-09-077-574A-1 (1-1647) x CH60_BARHE (1-547)

QY 1 ATGGCTTCTAAAGAAATCTTTTGTGCTAAAGCCGCGTAAAGAACTTTTACGAGGTGA 60

DB 1 MetAlaAlaLysGluValLysPheGlyAArgGluArgLeuArgGlyVal 20

QY 61 GATAAATCTGCAATCTGTAAAGTAACACTTGAACCTTAAAGCCGCTAATGCTGTTAT 120

DB 21 AspIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValle 40

QY 121 GAAAGCTTTTGGTTCCTCCAGTATTACAAAGATGGTGTATCTGTTCGCAAGAAAT 180

DB 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

QY 181 GAACCTGAAGATAGTTGAAATATGGCGCTCAATAGTTAAAGATAGCTCCCAAA 240

DB 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80

QY 241 ACTAGCGATATCTGCTGATGAAGTACAAACAGCAACAGTCTTCCACAAAGCTATTAT 300

DB 81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuGlyGlnAlaIleVal 100

QY 301 CGTGAAGGTGTAAGAACTGTAGCAGCTGGTGTATCTTATCTTATGCGCAATTAACGTGCATA 360

DB 101 GluGluGlyValLysAlaValAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

QY 361 GATAAAGCTGTTGTTGCTTACTAAAGAACTAAGACGACATTACAAAGCCTACTCTGAC 420

DB 121 AspAlaAlaValAspGluValValAlaAsnLeuPheLysLysAlaLysLysIleGlnThr 140

QY 421 CAAAAGAAATAGCTCAAGTTGGAACCACTTCTGCAAACTCTGTATACAACTAGTAAT 480

DB 141 SerAlaGluIleAlaGlnValGlyThrIleSerAlaAsnGlyAlaAlaGluIleGlyLys 160

QY 481 ATCATAGCTGAAGCTATGGCTTAAAGTTGGAAGAGGAGGTGTTATCACAGTTGAGAGCT 540

DB 161 MetIleAlaAspAlaMetGluLysValGlyAsnGluGlyValIleThrValGluAla 180

QY 541 AAAGCTCTTGAACACTACATAGCTGTTGAAGCAATGAAGTTTGACCGTGGCTACCTC 600

DB 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyIleLeu 200

QY 601 TCTCCATATCTTGAATCTCCTGAGAAATGGTTTGTGAACCTGTGATACCTTATATC 660

DB 201 SerProTyrPheValThrAsnAlaGluLysMetValAlaAspLeuAspAspProTyrIle 220

QY 661 CTTTGTATGCAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720

DB 221 LeuIleHisGluLysLysSerAsnLeuGlnSerLeuLeuProValLeuGluAlaVal 240

QY 721 GCTAAAGTAAACCGCTCCTCTTATTATTGCTGAAGACCTAGAAGCTGAAGCATTGCA 780

DB 241 ValGlnSerGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260

QY 781 ACATTTGTAGTCAATAAGCTCCGTCGAGCAGCTCCAAAGTTGTAGCCGTAAAGCTCTG 840

DB 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

QY 841 TTTGTGACGCGGTAAAGCTATGCTTGAAGATATTGCTATCTTACTGAGAGAGCA 900

DB 281 PheGlyAspArgLysAlaMetLeuGluAspIleAlaIleLeuThrSerGlyGlnVal 300

QY 901 ATATTGGAAGATCGGTATAAAGCTTGAAGATGTAAGCTTCTTCTTTAGCAACAGCT 960

DB 301 IleSerGluAspValGlyIleLysLeuGluAsnValThrLeuAspMetLeuGlyArgAla 320

QY 961 AAACGTGTAGTTATTGACAAAGAAATACTATCTGTTGTTGTTGTTGTTGTTGTTGTT 1020

DB 321 LysLysValAsnIleSerLysGluAsnThrThrIleIleAspGlyAlaGlyLysSer 340

QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGACAAACAGCTCAGATTAT 1080

DB 341 GluIleAsnAlaArgValAsnGlnIleLysValGlnIleGluIleThrThrSerAspTyr 360

QY 1081 GATCGTGAAGAACTTCAAGAACTCTTGCACAACTTGTGTTGAGTAGTGTATATCCAT 1140

DB 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

QY 1141 GTTGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

DB 381 ValGlyGlyAlaThrGluValGluValLysLysLysLysLysLysLysLysLysLys 400

QY 1201 AATGCAACAGAGCTCGGCTTGAAGAGGATTTGCTCCCTGTTGTTGTTGTTGTTGTTG 1260

DB 401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyThrAlaLeuLeu 420

QY 1261 CGTCCATTTAAAGTCTTATGATGATTAACCTGCTGATGATGATGATGATGATGATGAT 1320

DB 421 ArgAlaAlaAsnAlaLeu---ThrValLysGlySerAsnProAspGlnGluAlaGlyIle 439

QY 1321 ATATCATCTCGCTGTTCTTCTTGAAGAGCCTTTTACGTAATTCGTAATTCGTAATTCG 1380

DB 440 AsnIleValArgAlaAlaLeuGlnAlaProAlaArgGlnIleAlaThrAsnAlaGlyGlu 459

QY 1381 GAAGGTTCTATTGTTCTAGAAAAAGTTCTGTGAACCAAAA---CATGGTTTGGATTAAAT 1437

DB 460 GluAlaAlaIleValGlyLysValLeuGluAsnAlaAspThrPheGlyTyrAsn 479


```
QY 721 GCTAAAGTAAACCGTCCACCTCTTATTATTGCTGAAGACGTAGAGGTGAAGCACTTGCA 780
Db .....
QY 241 ValGlnThrGlyLysProLeuLeuLeuLeuLeuLeuLeuValGluGlyGluAlaLeuAla 260
Db .....
QY 781 ACACCTTGAGTCAATAAGCTCCGTGGACACTCCAAGTTGTAGCGTAAAGCTCCCTGGT 840
Db .....
QY 841 TTTGGTGAACCCGTAAGCTATGCTTCAAGATATTGCTATCCCTTACTTGGGGAGACCA 900
Db .....
QY 281 PheGlyAspArgArgLysAlaMetLeuGluAspLeuAlaLeuThrGlyGlyThrVal 300
QY 901 ATATTGAGATCGTGGTATTAAGCTTCAAAATGTAAGCTTCTCTTTAGGAACAGCT 960
Db .....
QY 301 IleSerGluAspLeuGlyLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320
QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCGTTGTAGTGGTGGGAAATCAGAA 1020
Db .....
QY 321 LysLysValSerIleThrLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340
QY 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATGGAACACAGCTCAGATTAT 1080
Db .....
QY 341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTGAAGAACTTCAAGAACTCTTCAAACTTTGTTGGTGGAGTAGCTGTATCCAT 1140
Db .....
QY 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleArg 380
QY 1141 GTTGAGCTGCTACTGAACTGAAATGAAAGAGAAAGAGATCGTGTGAGAGATGCTCTA 1200
Db .....
QY 381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgIleAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAAGAGTATTCCTCGTGGTGGTACTGCTTTTGTCT 1260
Db .....
QY 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeu 420
QY 1261 CGC---TCCATTAAAGTCTTATGATATTAACCTGCTGATGATGATGATGATGATGAT 1317
Db .....
QY 421 ArgSerValLysIle-----ThrValLysGlyLysGluAsnAspGlnAspAlaGly 438
QY 1318 CTTAATATCATCCGCTGCTTCTTGAAGAGCTTTTACGTCAAAATGCTGCAAAATGCTGCG 1377
Db .....
QY 439 ValAsnIleValArgAlaLeuGlnSerProAlaArgGlnIleValGluAsnAlaGly 458
QY 1378 TATGAGGTTCTATTGTTGTAAGAAAGTTCGTGACCAAAA---GATGTTTTCGATT 1434
Db .....
QY 459 AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspPheGlyTyr 478
QY 1435 AATGCTGCATCAGGAGATATGAAGACCTTATTAAGCTGCTGCTCATTCATCTAAAGAA 1494
Db .....
QY 479 AsnAlaGlnThrGlyGlyTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys 498
QY 1495 GTTACACGTTATGCTTACAAATACAGATCAGTACGCTCCTTACTTCTTAACTCAGAA 1554
Db .....
QY 499 ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu 518
QY 1555 TCGGCTATTGCTGAAACACCAAGCTTAAAGATATCCCTATGCTCGCGGTGTATG 1614
Db .....
QY 519 AlaMetIleAlaGluLeu-----ProLysLysAspAlaProAlaMetProGlyGlyMet 536
QY 1615 GTGGTATCGGTGGTATGACGATG 1641
Db .....
QY 537 GlyGlyMetGlyGlyMetAspMetMet 545
Db .....
RESULT 10
ID CH60 AMOPS STANDARD; PRT; 551 AA.
AC P26004;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR MOPA.
```

```
OS Amoeba proteus symbiotic bacterium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae.
OX NCBI_TaxID=2728;
RN [1]
RP SEQUENCE FROM N.A.
RA Ahn T.I., Leeu H.K., Kwak I.H., Jeon K.W.;
RT "Nucleotide sequence and temperature-dependent expression of groEL
gene isolated from symbiotic bacteria of Amoeba proteus.";
RL Endocyt. Cell Res. 8:33-44(1991).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M86549; AAC09381.1; -.
DR PIR; JC2562; JC2562.
DR HSP; P06139; IGR1.
DR HAMAP; MF_00600; -.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 551 AA; 58332 MW; A25895CBF9585DA0 CRC64;
Alignment Scores:
Pred. No.: 5,928-96 Length: 551
Score: 1866.00 Matches: 373
Percent Similarity: 81.42% Conservative: 74
Best Local Similarity: 67.94% Mismatches: 96
Query Match: 62.87% Indels: 6
DB: 1 Gaps: 2
US-09-077-574A-1 (1-1647) x CH60_AMOPS (1-551)
QY 7 TCTAAAGAAATCTTTTGTGATCTAAAGCCCGTGAAGAACTTTCACAGGTGTAGATAA 66
Db .....
QY 2 AlalysGluLeuArgPheGlyAspAlaArgGlnMetLeuAlaGlyValAsnAla 21
QY 67 CTTGCAAAATGCTGTTAAAGTAACTGACCTGACCTAAAGCCGTAATGTCCTATTGAAAG 126
Db .....
QY 22 LeuAlaAspArgValLysAlaThrMetGlyProSerGlyArgAsnValValLeuGluArg 41
QY 127 TCTTTTGTTCCTCCAGTATTACAAAGATGCTGTATCTGTTGCAAGAAATTTGAACCT 186
Db .....
QY 42 SerPheGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61
QY 187 GAAGATAAGTTTGAAATATGGCGCTCAAATGGTTTAAAGAGTAGTCCCAAACTAGC 246
Db .....
QY 62 GluAsnArgPheLysAsnMetGlyAlaGlnMetValLysGluValAlaLysThrSer 81
QY 247 GATATTGCTGTGATGGAACACTACACAGCAACAGCTCTTGCACACAGCTATTATCTGGA 306
Db .....
QY 82 AspThrAlaGlyAspGlyThrThrAlaThrValLeuAlaArgSerIleValValGlu 101
QY 307 GGTGTAAGAACTTGTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366
Db .....
QY 102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121
Db .....
```


CC -- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE007230; AAK5053.1; -;
 DR PIR; C95311; C95311.
 DR HAMAP; MF_00600; -; 1.
 DR InterPro; IPR001844; Chaperin_Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR Pfam; PF00118; cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 DR Chaperone; ATP-binding; Heat shock; Multigene family; Plasmid;
 KW Complete proteome.
 SQ SEQUENCE 545 AA; 57701 MW; 49BA61F23CE70431 CRC64;

Alignment Scores:
 Pred. No.: 7.62e-96 Length: 545
 Score: 1864.00 Matches: 387
 Percent Similarity: 81.24% Conservative: 59
 Best Local Similarity: 70.49% Mismatches: 97
 Query Match: 62.80% Indels: 6
 DB: 1 Gaps: 4

US-09-077-574A-1 (1-1647) x CH64_RHIME (1-545)

QY 1 ATGCTTCTAAAGAAATCTTTTGTATGCTTAAGCCCGTGAAGAACTTTCACGAGGTGA 60
 Db 1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuArgGlyVal 20
 QY 61 GATAAATCTGCAATGCTGTATAAGTAACTTCGACCTAAAGCCGTAATGCTGTTATT 120
 Db 21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValle 40
 QY 121 GAAAGCTTTTGTGTCCTCCAGTTATTACAAAGATGCTGATCTGTGTCGAAAGAAAT 180
 Db 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluile 60
 QY 181 GAATCTCAAGATAGTTGAAAATATGGCGCTCAATGTTAAAGATAGCTCCCAAA 240
 Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80
 QY 241 ACTAGCGATATTGCTGCTGATGGAACCTACACAGCAACAGTCCTTGACCAAGCTATTAT 300
 Db 81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaValle 100
 QY 301 CGTGAAGGTGTAACCTTTAGCAGCTGGTGGTAAATCTTATGCGCATTAACCGTGGCATA 360
 Db 101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLysArgGlyile 120
 QY 361 GATAAAGCTTGTGCTGTTACTAAAGAACTAAGCCACATTACAGCCTACTCGTGCAC 420
 Db 121 AspLeuAlaValAlaGluValValLysAspLeuAlaLysLysLysIleAsnThr 140
 QY 421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACCAAAATAGGTAAT 480
 Db 141 SerAspGluValAlaGlnValGlyThrIleSerAlaAsnGlyLysGlnIleGlyLeu 160
 QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAGAGGAGGTGTTATCAAGTTGAGGAAGCT 540
 Db 161 AspIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180
 QY 541 AAAGCTTTGAAACTACATTAGATGCTGTTGAAGGAATGAAGTTGACCGTGGCTACCTC 600
 Db 181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspA-gGlyTyrLeu 200

QY 601 TCTCATACTTTGTAACATACTCTGAGAAATGTTTGTGAACCTTGATACCCCTTATATC 660
 Db 201 SerProLysPheValThrAsnProLysMetValAlaAspLeuGluAspAlaPhele 220
 QY 661 CTTTGTATAGAAAGATTTACTAGCATGAAAGACATGCTACCAATCTTGAACCAAGTT 720
 Db 221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
 QY 721 CTTAAAGTAAACCGTCCACTCTTATTATTCTGAAGACGTTAGAGGTGAAGCACTTGA 780
 Db 241 ValGlnThrGlyLysProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
 QY 781 ACATTTGTAGTCAATAAGCTCCGTTGGAGCATCTCCAAGTTGTAGCGTGAAGCTCCTCGT 840
 Db 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280
 QY 841 TTTGCTGAACCCGCTAAAGCTTATGCTGAAGATATTGCTATCTTCTTACTGGAGGAGCA 900
 Db 281 PheGlyAspArgGlyAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300
 QY 901 ATATTGAGATCGTGGTATAAAGCTTGAAGTGAAGTCTTCTTCTTATAGAACAGCT 960
 Db 301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320
 QY 961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCGTTGATGCTGCTGGAATACGAA 1020
 Db 321 LysLysValSerIleThrLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340
 QY 1021 GATATTAAAGCTCGAGTTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTAT 1080
 Db 341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluLeuThrThrSerAspTyr 360
 QY 1081 GATCGTGAAGAACTTCAAGAACGCTTCGCAAACTTGTGTTGTTGAGTAGCTGTATTCAT 1140
 Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380
 QY 1141 GTTCGAGCTGCTACTGAAACTGAAATGAAAGAGAGAGAGATCTGCTGCTGCTGCTGCT 1200
 Db 381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400
 QY 1201 AATGCAACAAAGAGCTGCGGTTGAAGAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420
 QY 1261 CGC---TCCATTAAAGTCTTGTATGATATTAAACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
 Db 421 ArgSerValLysIle---ThrValLysGlyGluAsnAspAspGlnAspAlaGly 438
 QY 1318 CTTAATATCATCGTCTGCTCTTGAAGAGCCTTTAGCTCAAAATGCTGCAAAATGCTGCTG 1377
 Db 439 ValAsnIleValArgArgAlaLeuGlnSerProAlaArgGlnIleValGluAsnAlaGly 458
 QY 1378 TATGAAGTTTCTATTGTTGTAGAAAAGTTTCGTGAACCAAAA---GATGTTTGTGATTT 1434
 Db 459 AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspAspPheGlyTyr 478
 QY 1435 AATGCTGCATCAGAGATATGAGACCTTATTAAAGCTGGTGTATGATGCTTAAATAAAA 1494
 Db 479 AsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys 498
 QY 1495 GTTACAGTATTTCATTACAAATGAGCATCAGTACCTCTTCTTCTTCTTCTTCTTCTTCT 1554
 Db 499 ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu 518
 QY 1555 TGGCTATTCTCTGCAAAACCCAGAACCTTAAAGAAATATGCTATGCTGCTGCTGCTGCTG 1614
 Db 519 AlaMetIleAlaGluLeu-----ProLysLysAspAlaProAlaMetProGlyGlyMet 536
 QY 1615 GGTGGTATGGGTGATGACCGTATG 1641
 Db 537 GlyGlyMetGlyGlyMetAspMetMet 545

RESULT 13

CH60_XYLFA

ID CH60_XYLFA STANDARD; PRT; 547 AA.

AC Q9PFF2;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 60 KDa chaperonin (Protein Cpn60) (groEL protein).

GN GROEL OR GROEL OR XF0615.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI_TaxID=2371;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9a5c;

RA MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brionese M.R.S.,

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neco C.H.,

RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,

RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

RA Ho P.L., Honeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kurane E.E., Laigret F., Lambais M.R., Leite L.C.C.,

RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado A.M., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,

RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).

CC -!- FUNCTION: Prevents misfolding and promotes the refolding and

CC proper assembly of unfolded polypeptides generated under stress

CC conditions (By similarity).

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

CC 7 subunits (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; AE003907; AAF83425.1; -

DR PIR; F82783; F82783.

DR HSP; P06139; IGLR.

DR HAMAP; MF_00600; -; 1.

DR InterPro; IPR001844; Chaperonin Cpn60.

DR InterPro; IPR002423; Cpn60/TCF-1.

DR Pfam; PF00118; cpn60_TCF1; 1.

DR PRINTS; PR00298; CHAPERONIN60.

DR PRINTS; PR00304; TCOMPLEXTCF1.

DR PROSITE; PS00296; CHAPERONIN60; 1.

KW Chaperone; ATP-binding; Complete proteome.

SQ SEQUENCE 547 AA; 5757 MW; 30BE7F937CA7A9D8 CRC64;

```
QY 961 AAACGTGTAGTATTATGACAAAGAAATACTACTATCGTTGATGCTGCTGCAAAATCAGAA 1020
Dd 321 LysLysValArgValSerLysGluAsnThrThrIleLeuAspGlyIleGlyAspAsnAsp 340
QY 1021 GATATTAAAGCTCGAGTTAAACAATTCGTCGACAAATTCGTCGACAAAGCAAGCTCAGATT 1080
Dd 341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360
QY 1081 GATCGTCAAAACCTTCAAGAACGCTTTCGAAACCTGTTGCTGAGTAGCTGTTATCCAT 1140
Dd 361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380
QY 1141 GTTGGAGCTGCTACTGAACTGAATGAAGAGAGAGAGAGAGAGATCGTAGAGAGCTGCTA 1200
Dd 381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAspAlaLeu 400
QY 1201 AATGCAACAGAGCTGGCTGGAAGAGGTAATTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Dd 401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420
QY 1261 CGCTCCATTAAAGCCCTTGATGATTAAACCTGCTGATGATGATGATGATGATGATGAT 1320
Dd 421 ArgAlaIleThrAlaIleSerAsnLysLysGlyAlaAsnGluAspGlnThrHisGlyIle 440
QY 1321 AATATCATCGCTGCTCTCTTGAAGAGCTTTAGCTCAAAATTCGTCGAAATGCTGCTGCTAT 1380
Dd 441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460
QY 1381 GAAGGTCTATTGTTGAGAAAAGTTCGTAACCAAGATGCTGTTTGGATTAAATGCT 1440
Dd 461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnProGlyTyrAsnAla 480
QY 1441 GCATCAGGAGAAATGAAGACCTTATTAAAGCTGCTGATGATGATGATGATGATGATGAT 1500
Dd 481 AlaThrGlyGluPheGlyAspMetValAsnLeuGlyIleLeuAspProThrLysValThr 500
QY 1501 CGTATTGCAATTAACAAATGAGCATGATGAGCTGCTTACTTCTTAACCTACAGATGCGCT 1560
Dd 501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520
QY 1561 ATTGCTGAAAACACAGAACTTAAAGATATATGCTATGCTGCTGCTGCTGCTGCTGCTGCT 1611
Dd 521 ValAlaGlu-----AlaProLysLysAspGluProThrProAlaAlaGlyGlyGly 538
QY 1612 ATGGGTGGTATGGTGGTATGGAC 1635
Dd 539 MetGlyGlyMetGlyGlyMetAsp 546
```

RESULT 14

```
CH60 RHOMR
ID -CH60 RHOMR STANDARD; PRT; 540 AA.
AC Q9XCA9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL.
OS Rhodothermus marinus (Rhodothermus obamensis).
OC Bacteria; Bacteroidetes; Spingobacteria; Spingobacteriales;
OC Crenitrichaceae; Rhodothermus.
OX NCBI_TaxID=29549;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=ITI 376;
RA Thorolfsdottir E.T.T., Backman V.M., Blondal T.,
RA Thorbjarnardottir S.H., Palsdottir A., Hauksdottir H.,
RA Kristjansdottir S., Eggertsson G.;
RT "Heat shock in Rhodothermus marinus: cloning and sequence analysis of
RT the groEL, dnaK and dnaJ genes."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
```

```
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF145252; AAD37976.1; -.
CC HSP; P06139; 1GRL.
CC HAMAP; MF_00600; -.
CC InterPro; IPR001844; Chaperonin Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC Pfam; PF00118; cpn60_TCF1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCPI.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC SEQUENCE 540 AA; 57666 MW; 686744837BEP18EC CRC64;
Alignment Scores:
Pred. No.: 3,28e-95 Length: 540
Score: 1852.50 Matches: 371
Percent Similarity: 82.46% Conservative: 71
Best Local Similarity: 69.22% Mismatches: 93
Query Match: 62.42% Indels: 1
DB: 1 Gaps: 1
```

US-09-077-574a-1 (1-1647) x CH60_RHOMR (1-540)

```
QY 1 ATGCTCTTAAGAAATCCCTTTTGTGCTTAAAGCCGCTGAAAACTTTCACGAGGTGA 60
Dd 1 MetAlaAlaLysGlnIleThrPheAsnAlaAspAlaArgMetAlaLeuLysArgGlyVal 20
QY 61 GATAAACTTCGCAATGCTGTTAAAGTAACACTTGGACCTTAAAGCCGCTTAATGCTGTTAT 120
Dd 21 AspLysLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValIle 40
QY 121 GAAAAGCTTTTGGTCCCGCTTATACAAAAGATGGTGTATCTGTGTCGCAAGAAAT 180
Dd 41 GluLysLysPheGlyAlaProThrValThrLysAspGlyValThrValAlaLysGluIle 60
QY 181 GAACTTGAAGATAAGTTGAAAATATGGCGCTCAATGCTTAAAGAGTAGCTCCCAA 240
Dd 61 GluLeuGluAspLysLeuGluAsnValGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTCCTGCTGATGGAACCTACAAACAGCAACAGCTCTTCGCAAGCTATTAT 300
Dd 81 ThrSerAspValAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleLeu 100
QY 301 CGTGAAGGTGTAACCTTGTAGCAGCTGTCGCTGCTTCTTATGCTGCTGCTGCTGCTGCTA 360
Dd 101 ThrAlaGlyLeuLysSerValThrAlaGlyAlaAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAGCTGTTGCTGCTGCTTAAAGAACTAAGCGACATACAAAGCTCTCTGCTGAC 420
Dd 121 AspLysAlaValGluValValValAlaGluLeuArgLysMetSerGlnGluValGlnAsp 140
QY 421 CAAGAAGAAATAGCTCAAGCTTGAACCTTCTGCAAACTCTGATACACAACTAGTAAAT 480
Dd 141 LysAsnArgIleAlaGlnValAlaThrIleSerAlaAsnGlyAspLysAlaIleGlyGln 160
QY 481 ATCATAGCTGAAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Dd 161 LeuIleAlaAspAlaPheGluLysValGlyLysAspGlyValIleThrValGluAla 180
QY 541 AAAGGTCTTGAACCTACATTAGTGTGTTGAAGGATGAAGTGTGACCGTGGCTACCTC 600
```



```
Db 181 LysGlyThrGluThrThrLeuGluValValGluGlyMetGlnPheAspArgGlyLeu 200
Qy 601 TCCTCCACTTTGTAACCTGCTGAGAAAGTGTGTAAGTGTGTAACCTGTAATATC 660
Db 201 SerProTyrPheValThrAsnProAspThrMetGluAlaValLeuGluAspAlaTyrIle 220
Qy 661 CTTTGTATGAGAAAGATTAAGTCTAGCATGAAGCATGCTACCAATCTTACGACAGTT 720
Db 221 LeuIleHisAspLysLysIleSerAlaMetLysAspLeuLeuProIleLeuGluLysVal 240
Qy 721 GCTAAGTAACCGCTCCACTCTTATTATGCTGAAGACGTAGAGGTGCAAGCTTGCA 780
Db 241 ValGlnThrGlyArgProLeuLeuIleAlaGluAspValGluGlyGluAlaLeuAla 260
Qy 781 ACATTGTAAGTCAATTAAGTCTCGTGGAGCATCTCAAGTGTGAGCGTAAAGCTCTGCT 840
Db 261 ThrLeuValValAsnLysLeuArgGlyValLeuLysValAlaAlaValLysAlaProGly 280
Qy 841 TTTGTGTAACCGCTGTAAGCTATGCTGAAGATATTGCTATCTTACTGAGGAGAGCA 900
Db 281 PheGlyAspArgGlyAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300
Qy 901 ATATTGTAAGTCTGCTGTAAGCTTGAAGCTTGAAGCTTCTTCTTTAGGAACAGCT 960
Db 301 IleSerGluGluLysGlyTyrArgLeuGluAsnAlaThrLeuAspTyrLeuGlyGlnAla 320
Qy 961 AAACGTGTAGTATTACAAAGAAATACTATCTATCTGTTGAGTGTGCGTGAAGTCA 1020
Db 321 GluArgIleIleValAspLysAspAsnThrThrIleValGlyGlyLysGlyAspProAla 340
Qy 1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAAGAACAGCTCAGATT 1080
Db 341 GlnIleLysAlaArgAlaAsnGlnIleArgGlnGlnIleGluGluThrThrSerAspTyr 360
Qy 1081 GATCGTGAAGAACTTCAAGAACTCTTCAAGAACTTGTGAGAGTACTGTTTCCAT 1140
Db 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValLeuLys 380
Qy 1141 GTTGAGCTGCTACTGAACTGAATGAAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
Db 381 IleGlyAlaAlaThrGluProGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
Qy 1201 AATGCAAGAGCTCGCTGTAAGAGTATGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaTyrIle 420
Qy 1261 CGTCTCATTAAAGTCTGATGATATTAACCTGCTGATGATGATGATGATGATGAT 1320
Db 421 ArgAlaIleAlaAlaLeuAspLysValGluValGluAsnGluAspGlnLysIleGlyVal 440
Qy 1321 AATATCATCGCTGCTCTTGAAGAGCTTTACGTCAATATGCTCAATATGCTGCTAT 1380
Db 441 GlnIleValGlnArgAlaLeuGluGluProLeuArgGlnIleAlaAlaAsnAlaGlyTyr 460
Qy 1381 GAAGTCTTATGCTGTAGAAAAGTCTGTAACCAAGATGTTTGGATTTAATGCT 1440
Db 461 GluGlySerIleValValGlnArgValLysGluGlyGlnGlyAspPheGlyPheAsnAla 480
Qy 1441 GCATCAGGAAATATGAAGACCTTATTAAAGTGTGCTCATTCATTCCTAAAAAGTTACA 1500
Db 481 GlnThrGluGluPheGlyAsnLeuLeuGluGlnGlyValIleAspProThrLysValAla 500
Qy 1501 CGTATTGCAATCAAAATGAGATCAGTACGCTCTTACTTCTTACTACTAGATGCGCT 1560
Db 501 ArgThrAlaLeuGluAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrGluAlaVal 520
Qy 1561 ATTGCTGAAAACACAGAACCTTAAAAA---GATATGCTATGCTGCTGCTGCTGCT 1605
Db 521 ValAlaGluLysProGluLysGluLysAlaAlaProSerProGly 536
```

RESULT 15
CH60_CHRVI

```
ID CH60_CHRVI STANDARD; PRT; 546 AA.
AC P31293;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL OR MOEA.
OS Chromatium vinosum.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
OC Chromatiaceae; Allochrochromatium.
OX NCBI_TaxID=1049;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93186721; PubMed=8444812;
RA Ferreyra R., Soncini F., Viale A.M.;
RT "Cloning, characterization, and functional expression in Escherichia
RT coli of chaperonin (groESL) genes from the phototrophic sulfur
RT bacterium Chromatium vinosum.";
RL J. Bacteriol. 175:1514-1523(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99008942; PubMed=9790891;
RA Dionisi H.M., Viale A.M.;
RT "Purification and characterization of Chromatium vinosum GroEL and
RT GroES proteins overexpressed in Escherichia coli cells lacking the
RT endogenous groESL operon.";
RL Protein Expr. Purif. 14:275-282(1998).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions.
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M99443; AAA23319.1; -
CC PIR; B47073; B47073.
CC HSP; P06139; 1GRL.
CC HAMAP; MF_00600; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chapterone; ATP-binding.
CC SEQUENCE 546 AA; 57541 MW; 3250141881C04DD6 CRC64;

```

Alignment Scores:
Pred. No.: 3-28e-95 Length: 546
Score: 1852.50 Matches: 374
Percent Similarity: 81.02% Conservative: 70
Best Local Similarity: 68.25% Mismatches: 101
Query Match: 62.42% Indels: 3
DB: 1 Gaps: 2

US-09-077-574A-1 (1-1647) x CH60_CHRVI (1-546)

```
Qy 1 ATCGCTTCTAAAGAAATCCTTTTGTATGCTAAAGCCGCTGAAACATTTTCAGAGGTCTA 60
Db 1 MetSerAlaLysAspValLysPheGlyGlyAspAlaArgValArgMetMetGluGlyVal 20
Qy 61 GATAAACTGCAATCTGTTAAAGTAACACTTGGACCTAAAGCCGCTTAATGCTGTTATT 120
Db 21 AsnIleuAlaAsnAlaValThrLeuGlyProLysGlyArgAsnValValLeu 40
```


[illegible]

Qy	1201	AATGCACACAGAGCTGGCGTTGAAAGAGTATTGTCCTCGTGGTGTACTGCTTTGTCT	1260
Db	401	HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyValAlaLeuVal	420
Qy	1261	CGTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGAACCTTGCTGGAATT	1320
Db	421	ArgAlaIleAlaValLysAspLeuLysGlyAlaAsnHisAspGlnAspValGlyIle	440
Qy	1321	AATATCATCCGTCGTTCTCTTGAAGAGCCTTTTACGTCAAAATGCTGCAAAATGCGGCTAT	1380
Db	441	AlaIleAlaArgAlaMetGluGluProLeuArgGlnIleValAlaAsnAlaGlyGlu	460
Qy	1381	GAGGTTCTATTGTTGAGAAAAAGTTCGTGAACCAACCAAGATGGTTTTGGATTAAATGCT	1440
Db	461	GluProSerValIleLeuHisLysValAlaGluGlyThrGlyAsnPheGlyTyrAsnAla	480
Qy	1441	GCATCAGGAGATATGAAGACCTTTATTAAAGCTGGTGTCTATTCATTCCTAAAAAGTTACA	1500
Db	481	AlaAsnGlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThr	500
Qy	1501	CGTATTGCATTACAAATGCAGACATCAGTAGCTCTCTACTTCTAACTACAGAAATGGCT	1560
Db	501	ArgSerAlaLeuGlnAsnSerCysSerValAlaGlyLeuMetIleThrThrGluAlaMet	520
Qy	1561	ATTGCTGAAAACCACCAACCTATAAAAGATATGCCCT---ATGCCTGGCGGTGTATGGGT	1617
Db	521	IleAlaasp-----GluProLysAspAspAlaProAlaMetProGlyGlyGlyMetGly	538
Qy	1618	GGTATGGTGGTATGACCGGTATG	1641
Db	539	AspMetGlyGlyMetGlyMetMet	546

Search completed: January 28, 2004, 13:26:11
Job time : 45.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:21:36 ; Search time 72 seconds
(without alignments)
11805.907 Million cell updates/sec

Title: US-09-077-574A-1

Perfect score: 2968

Sequence: 1 agggctcaagaatactct.....gtatgacggtatgtactag 1647

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool_p/US09077574/runat 28012004 125846 11617/app query.fasta_1.1799
-DB=SPTRMBL 23 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEA SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574 -CGEN 1 1 92 @runat 28012004 125846 11617 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6
-PGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2708	91.2	548	2 O87888

ALIGNMENTS

RESULT 1

ID	O87888	PRELIMINARY;	PRT;	548 AA.
AC	O87888;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	GroEL/HSP60 homolog (60 kDa chaperonin) (Protein Cpn60) (groEL protein).			
DE	Lawsonia intracellularis.			
OG	Plasmid pISI-2.			
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfotribionales;			
OC	Desulfotribionaceae; Lawsonia.			
OX	NCBI_TaxID=29546;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98386497; PubMed=9720028;			
RA	Dale C.J.H., Moses E.K., Ong C.C., Morrow C.J., Reed M.B., Hasse D.,			
RA	Strugnelli R.A.;			
RT	"Identification and sequencing of the groE operon and flanking genes of Lawsonia intracellularis: use in phylogeny.";			
RL	Microbiology 144:0-0(0).			
CC	-I- FUNCTION: PREVENTS FOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).			

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OX Acetobacteraceae; Acetobacter.

RN NCBI_TaxID=435;

RA [1]

SEQUENCE FROM N.A.

RC STRAIN=IFO3283;

RA Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,

RA Yanagida F.;

RT "Cloning and Characterization of groEL Operon in Acetobacter aceti.";

RL J. Biosci. Bioeng. 94:140-147 (2002).

DR EMBL; AB081586; BAC16232.1; -

SQ SEQUENCE 546 AA; 58106 MW; 3E476AF42F2CD5CC CRC64;

Alignment Scores:

Pred. No.:	9,17e-91	Length:	546
Score:	1880.00	Matches:	381
Percent Similarity:	82.42%	Conservative:	69
Best Local Similarity:	69.78%	Mismatches:	94
Query Match:	63.34%	Indels:	2
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x Q8BBD2 (1-546)

```

QY 1 ATGGCTTCTAAAGAAATCCCTTTTGATGCTAAAGCCCGTGAAACATTTCCAGAGGTGA 60
DB 1 MetAlaAlaLysAspValLysPheGlyAlaAspAlaArgGlnArgMetArgAlaGlyVal 20
QY 61 GATAAATCTGCAATCTGTTAAAGTAAACCTTGGACCTTAAGCCGCTAATGCTGTTAT 120
DB 21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGCTCTTTGGTCCCGCAGTTATTACAAAAGATGGTGTATCTGTGTCGAAAGAAAT 180
DB 41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60
QY 181 GAACCTTCAAGATAAAGTTGAAATATGGCGCTCAAAATGGTTAAAGAGTAGTCCCAA 240
DB 61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80
QY 241 ACTAGCGATATCTCGTGTAGTGAACATACACAGACAGCTCTTGCACAGCTATTAT 300
DB 81 ThrAsnAspIleAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CGTGAAGGTGTAACACTTGTAGCAGCTGGTGGTAATCCCTATGCCATTAAACGTCGCA 360
DB 101 ArgGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTGTTGCTGCTGTTACTAAAGAACTAAGCGACATTAACAAAGCCCTACTCG 420
DB 121 AspLysAlaValAlaValValIleGluGluLeuLysLysAsnAlaLysLysValThrThr 140
QY 421 CAAAAGAAATAGCTCAAGTTGAAACCATTTCTGCAACTCTGCAACATCTGATCAACAT 480
DB 141 ProAlaGluThrAlaGlnValGlyThrIleSerAlaAsnGlyGluSerGluIleGlyGln 160
QY 481 ATCATAGCTGAAGCTATGGCTAAAGTTGAAAGAGGAGGTCTTATCAGTTGAGGAGCT 540
DB 161 MetIleSerGluAlaMetGlnLysValGlySerGluGlyValIleThrValGluGluAla 180
QY 541 AAAGGTCTTTGAAACTACATTAGATGTTGGTTGAAGGAATGAAGTTGACCGTGCTACCTC 600
DB 181 LysHisPheGlnThrGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrlle 200
QY 601 TCTCCATCTTTGTAACATCTCTGAGAAATGGTTTGCACTGATTAACCTTATATC 660
DB 201 SerProTyThrPheValThrAsnProGluLysMetThrAlaAspLeuGluAsnProTyrlle 220
QY 661 CTTTGTAACTGAGAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
DB 221 LeuIleHisGluLysLysLeuSerSerLeuGlnProMetLeuProLeuGluSerVal 240
QY 721 GCTAAAGTAAACCGTCCACTCTCTATTATTGCTTGAAGACGTAGAAGGTGAAGCATTGCA 780
DB 780

```

```

DB 241 ValGlnSerGlyArgProLeuLeuIleAlaGluAspValAspGlyGluAlaLeuAla 260
QY 781 ACACCTTGTAGTCAATAAGCTCCGTCGAGCACTCCAAGTTGTAGCCGCTAAAGCTCTGGT 840
DB 261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaValLysAlaProGly 280
QY 841 TTTGGTGAACCGCCCTAAAGCTATGCTTGAAGATATTGCTATCTTACTCGAGGAGAGCA 900
DB 281 PheGlyAspArgAsnValMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300
QY 901 ATATTTCAAGATCGTGTATATAAGCTTTGAAATGTAAAGCTTGTCTTCTTTAGGAACAG 960
DB 301 IleSerGluAspLeuGlyIleLysLeuGluThrValThrLeuAsnMetLeuGlyThrAla 320
QY 961 AAACGTGTAGTTATTGACAAAGAAATATCTACTATCTGTTGATGCTGCTGAAATCAGAA 1020
DB 321 LysLysValHisIleAspLysGluAsnThrThrIleValAspGlyAlaGlyLysAlaAsp 340
QY 1021 GATATTAAGCTCGAGTTAAACAAATTCGTGCACAAATTCGAAGAAACAAGCTCAGATTAT 1080
DB 341 AspIleLysGlyArgValLysGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360
QY 1081 GATCGTGAACAACTTCAAGAACCTTTCGCAAACTTGTGGTCGAGTAGCTGTATCCAT 1140
DB 361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaIleArg 380
QY 1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAAAGAGATCGTGTAGAGATGCTCTA 1200
DB 381 ValGlyGlySerThrGluValGluValLysGluArgLysAspArgValAspAlaLeu 400
QY 1201 AATGCAACAGAGCTCGGTTGAAGAGGTATTGTCCTGCTGCTGCTGCTGCTTTGTC 1260
DB 401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyThrAlaLeuAla 420
QY 1261 CGCTCCATTAAGTCCCTGATGATATTAAACCTGCTGATGATGATGATGCTGCTGCT 1320
DB 421 ArgAlaThrLeuLysLeuGluGlyLeuHisTyrHisAsnAspAspGlnArgValGlyGly 440
QY 1321 AATATCATCGCTGCTCTTGTGAAGAGCTTTAGTCAATTCGTCGCAATGCTGCTGCTAT 1380
DB 441 AspIleAlaGluAlaLeuGlnAlaProLeuArgGlnIleAlaHisAsnAlaGlyGlu 460
QY 1381 GAAGGTTCTATTCTGTAGAAAAGTTCGTGAACCAACAGAT--GTTTTGCTATTAT 1437
DB 461 AspGlyAlaValIleAlaAsnLysValLeuGluAsnSerAspTyrAsnPheGlyPheAsp 480
QY 1438 GCTGCATCAGGAATATGAAGACCTTATTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497
DB 481 AlaGlnAlaGlyGluTyrLysAsnLeuValGluAlaGlyIleIleAspProAlaLysVal 500
QY 1498 ACACGTATTGCATTACAAATGCAGCATCAGTAGCTCTCTTACTTCTAATCAGAAATGC 1557
DB 501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuIleThrThrGluAla 520
QY 1558 GCTATGCTGAAAACCCAGAACCTTAAAGAAATATGCTGCTGCTGCTGCTGCTGCTGCT 1617
DB 521 MetValAlaGluArgProGlu---LysLysAlaAlaProAlaGlyGlyProAspMetGly 539
QY 1618 GGTATGGGTGGTATGGAC 1635
DB 540 GlyMetGlyGlyMetAsp 545

```

RESULT 3

Q8KJ46

ID

Q8KJ46;

AC

DT

DT

DT

DE

DE

GN

OS

PRELIMINARY; PRT; 552 AA.

Q8KJ46; (TrEMBLrel. 22, Created)

01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

Probable chaperonin groEL DF protein (60 kDa chaperonin) (Protein

Cpn60) (groEL protein)

GN GROEL

OS Rhizobium loti (Mesorhizobium loti).

OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;	Db	161	MetileAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla	180
OC	Phyllobacteriaceae; Mesorhizobium.	Qy	541	AAAGTCTTTGAACACTACATTAGATGCTGCTGAAGGAATGAAGTTTGACCGTGGCTACCTC	600
OX	NCBI_TaxID=381;	Db	181	LysThrAlaGluThrGluGluValGluGlyMetGlnPheAspArgGlyTyrLeu	200
RN	[1]	Qy	601	TCTCATTAATTTGTAACCTCCTGAGAAAATGTTTGTGAACCTTGTATACCTTATATC	660
RC	SEQUENCE FROM N.A.	Db	201	SerProTyrPheValThrAsnAlaAspLysMetValAlaAspLeuGluAspValTyrIle	220
RX	MEDLINE=21999272; PubMed=12003951;	Qy	661	CTTTGTAATGAGAAAAGATTCTAGCATGAAAGACATCTACCAATCTTAGAACAGTT	720
RA	Sullivan J.T., Tzabiatowski J.R., Cruickshank R.W., Gouzy J.,	Db	221	LeuLeuHisGluLysLysLeuSerAsnLeuGlnThrMetLeuProValLeuGluAlaVal	240
RA	Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,	Qy	721	GCTAAGTAAACCGTCCACTCCTTATTATGCTGAAGACGTAGAAGGTGAACCTTGA	780
RA	Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;	Db	241	ValGlnThrSerLysProLeuLeuIleSerGluAspValGluGlyGluAlaLeuAla	260
RT	"Comparative sequence analysis of the symbiosis island of	Qy	781	ACACTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCGTGAAGCTCCTG	840
RT	Mesorhizobium loti strain R7A."	Db	261	ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly	280
RL	J. Bacteriol. 184:3086-3095(2002).	Qy	841	TTTGTGTAACCGCTTAAGCTATGCTTGAAGATTTGCTATCTTACTTGGAGGAGCA	900
CC	-!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND	Db	281	PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal	300
CC	CONDITIONS (BY SIMILARITY).	Qy	901	ATATTGTAAGATCGTGTATTAAGCTTGAAGATTAAGCTTCTTCTTTAGGAACAGCT	960
CC	-!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF	Db	301	IleSerGluAspLeuGlyLysLeuGluAsnValGlyLeuAsnMetLeuGlyArgAla	320
CC	7 SUBUNITS (BY SIMILARITY).	Qy	961	AAACGTGTAGTTTATTGACAAAATACTACTATCTGTTGATGCTGGTGGAAAATCAGAA	1020
CC	-!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.	Db	321	LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyLysLysGlu	340
DR	EMBL; AL672115; CAD31231.1; InterPro; IPR001844; Chaprinin_Cpn60.	Qy	1021	GATATTAAGCTCGATTAAACAAATTCGTGCACAAATTAAGAAACAGCTCAGATTAT	1080
DR	InterPro; IPR002423; Cpn60/TCP-1.	Db	341	GluIleGlnGlyArgValAlaGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr	360
DR	Pfam; PF00118; cpn60 TCP1; 1.	Qy	1081	GATCGTGAAGAACTTCAAGAACGCTTGCAAAACCTTGTGGTGGAGTAGCTGTATCCAT	1140
DR	PRINTS; PR00298; CHAPERONIN60.	Db	361	AspLysGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleArg	380
DR	PRINTS; PR00304; TCOMPLEXTCP1.	Qy	1141	GTTGAGCTGCTTACTGAAACTGAAATGAAAGAGAAGAGGATCGTGTAGAAGATGCTTA	1200
DR	PROSITE; PS00296; CHAPERONINS_CPN60; 1.	Db	381	ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAlaLeu	400
KW	ATP-binding; Chaperone.	Qy	1201	AATGCAACAGAGCTGCGTTGAAGAGTATTGTCCTCGTGGTGGTGTACTGCTTTGTC	1260
SQ	SEQUENCE 552 AA; 57747 MW; 7D0A8C53CD7FA6AB CRC64;	Db	401	AsnAlaThrArgAlaAlaValGluGlyIleValProGlyGlyGlyValAlaLeu	420
		Qy	1261	CGCTCCATTAAAGTCTTGTATGATATTAAACCTGCTGATGATGATGATGATGATGAT	1320
		Db	421	ArgAlaSerLeuSerIleAsnAlaVal---GlyAlaAsnSerAspGlnAlaGlyIle	439
		Qy	1321	AATATCATCCGCTGCTTCTTGAAGAGCCTTTTACGTCAAATTTGCTCAAATCTGCTGCTAT	1380
		Db	440	AsnIleValArgArgAlaLeuGlnAlaProAlaArgGlnIleAlaAlaAsnAlaGlyAla	459
		Qy	1381	GAAGTCTTATTGTTGTAGAAAAGTTCGTGACCAACCAAGATGGT---TTTGATTAT	1437
		Db	460	GluAlaSerIleValAlaGlyLysIleLeuGluAsnLysGlyAlaThrPheGlyTyrAsn	479
		Qy	1438	GCTGCATCAGGAGATATGAACACCTTATTAAAGCTGGTGTCTATTGATCTCTAAAAGTT	1497
		Db	480	AlaGlnThrGlyGlyLysGlyAspMetIleAlaMetGlyIleValAspProValLysVal	499
		Qy	1498	ACACGATTATGCTTACAAAATGCAGCATCAGTAGCTCTCTTACTTCTTAAGTACAGAAATGC	1557
		Db	500	ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla	519
		Qy	1558	GCTATGCTGTAAGAAACCAACCACTAAAAA-----GATATGCT-----ATG	1599
		Db	520	MetIleAlaGluAlaProLysLysGluSerAlaGlyGlyGlyMetProGlyGlyMet	539

Alignment Scores:

Pred. No.:	8.e-88	Length:	552
Score:	1823.00	Matches:	374
Percent Similarity:	80.98%	Conservative:	73
Best Local Similarity:	67.75%	Mismatches:	97
Query Match:	61.42%	Indels:	8
DB:	2	Gaps:	4

US-09-077-574A-1 (1-1647) x Q8KJ46 (1-552)

Qy	1	ATGCGTCTTAAGAAATCTTTTGTATGCTAAAGCCGCTGAAAACCTTTCACAGGTGTA	60
Db	1	MetAlaAlaLysAspValLysPheSerArgAspAlaArgGluArgMetLeuArgGlyVal	20
Qy	61	GATAAATGCAAACTGTTAAAGTAAACACTGACCTTAAAGCCGCTAAATCTGCTTAT	120
Db	21	AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValIle	40
Qy	121	GAAAGTCTTTGGTTCCTCCAGTATTACAAAGATGGTGTATCTGTTGCAAAAGAAAT	180
Db	41	AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle	60
Qy	181	GAACCTGAAGATAGTTTGAATAATATGGCGCTCAAAAGTGTAAAGAGTAGCTCCCAA	240
Db	61	GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys	80
Qy	241	ACTAGCATATGCTGGTGTAGTGAACCTACACAGCAACAGTCTCTTGCAAGCTATTAT	300
Db	81	ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal	100
Qy	301	CGTGAAGTGTAAAACCTTGTAGCAGTGTGCTTAATCTCTATGGCCATAAAGCTGGCATA	360
Db	101	GlnGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle	120
Qy	361	GATAAGCTGTTGTTGCTGTTACTAAGAACTAAGAGCAATACAAAGCTTACTCGTGAC	420
Db	121	AspLeuAlaValThrAspValValAlaThrLeuIleLysAsnAlaLysLysIleYthr	140
Qy	421	CAAAAGAAATAGCTCAAGTGTGAACCTTCTGCAAACTCTGATACACAAATAGTAAT	480
Db	141	SerGluGluValAlaGlnValGlyThrIleAlaGlyAsnGlyAspGluSerValGlyLys	160
Qy	481	ATCATAGCTGAAGCTATGCTAAAGTTGGAAGAGGAGGTGTTATCACAGTTGAGGAGCT	540


```
QY 1495 GTTACAGCTATTGATTAACAAATGACATCAGTAGCTCTTACTTCTTAACACAGAA 1554
Db 499 ValValArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuIleThrGlu 518
QY 1555 TGCCTATTGCTGAAAAACCAAGACCTTAAAAAGATATGCTATGCGCTGGC-----GGT 1608
Db 519 AlaMetIleAlaGluLeu-----ProLysLysAspAlaAlaProAlaGlyMetProGly 536
QY 1609 GGTATGGTGGTATGGGTGGTATGGAC 1635
Db 537 GlyMetGlyGlyMetGlyGlyMetAsp 545

RESULT 5
Q8GB95 PRELIMINARY; PRT; 545 AA.
AC Q8GB95;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 60 kDa chapertonin GroEL (Fragment).
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22337798; PubMed=12446909;
RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
RA Blankenship R.E.;
RT "Whole-genome analysis of photosynthetic prokaryotes.";
RL Science 298:1616-1620(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
RA Gerdes S., Kyrpides N., Overbeek R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY142910; AAN87514.1; -.
FT NON TER 545
SQ SEQUENCE 545 AA; 57927 MW; B9F6DB02DBC18383 CRC64;

Alignment Scores:
Pred. No.: 8,16e-87 Length: 545
Score: 1804.50 Matches: 360
Percent Similarity: 81.3% Conservative: 84
Best Local Similarity: 65.9% Mismatches: 99
Query Match: 60.8% Indels: 3
Gaps: 2
DB: 3

US-09-077-574A-1 (1-1647) x Q8GB95 (1-545)
QY 7 TCTAAGAAATCCTTTTGTATGCTAAAGCCGTGAAAACTTTCAGAGGTGTAGATAA 66
Db 2 AlaLysMetThrValPheAsnGluGluAlaArgAlaLeuGluLysGlyValAsnThr 21
QY 67 CTTGCAAAATGCTGTTAAAGTAACTTGGACCTAAAGGCGGTAATGCTGTTAAAGAAAG 126
Db 22 LeuAlaGluAlaValArgValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41
QY 127 TCTTTTGGTCCCGATTAATACAAAGATGTGTATCTGTGCAAAAGAAATGCACTT 186
Db 42 LysPheGlySerProLeuIleThrAsnAspGlyValThrIleAlaLysGluIleGluLeu 61
QY 187 GAAGATAAGTTTGAATAATGGCGCTCAATGGTTAAAGAGTAGTCCCAAAACTAGC 246
Db 62 GluAsnProIleGluAsnMetGlyAlaGlnLeuValLysGluValAlaThrLysThrAsn 81
QY 247 GATATTGCTGGTGGTAACTACACAGCAAGCTCTTGGCAAGCTATTATTCGTGAA 306
Db 82 AspValAlaGlyAspGlyThrThrAlaThrIleLeuAlaGlnAlaIleIleArgGlu 101
QY 307 GGTGTAAACTTGTAGCAGCTGCTGTATCTATGCGCATTTAAAGCTGCATAGATAA 366
Db 102 GlyMetLysAsnValAlaAlaGlyAlaAsnProMetValIleLysArgGlyIleGluLys 121
```

```
QY 367 GCTGTTGTTGCTTACTAAAGAACTAAGCGACATTAACAAAGCCCTACTCGTGACCAAAA 426
Db 122 AlaValAlaThrCysValGluGluIleLysAsnLeuSerLysProValGluSerLysGlu 141
QY 427 GAAATAGCTCAAGTTGGAAACCATTTCTGCAAACTCTGTATACAAATAGGTAAATCAT 486
Db 142 AlaIleAlaGlnValAlaSerIleSerAla---AlaAspAlaThrIleGlySerLeuIle 160
QY 487 GCTGAAGCTATGCTTAAAGTTGGAAAAGAGGTGTATACAGATTGAGAGCAAGCTAAGGT 546
Db 161 AlaGluAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluSerLysGly 180
QY 547 CTTGAAACTACATTAGATGTGTTGAAGCAATGAAGTTTGACCGTGGCTACCTCTCTCA 606
Db 181 PheSerThrAspLeuGluValValGluGlyMetAsnPheAspArgGlyTyrIleSerPro 200
QY 607 TACTTTGTAATATCTCTGAGAAATGTTTGTGAACCTGTATACCCCTTATATCTTTGT 666
Db 201 TyrMetIleThrAspProAspLysMetGluAlaValLeuAsnAspProTyrIleLeuIle 220
QY 667 AATGAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACCAAGCTTGCTAA 726
Db 221 ThrAspLysLysIleSerAlaIleLysAspIleLeuProValLeuGluArgIleValGln 240
QY 727 GTAAACCCCTCCACTCTTATTATCTGAAGACGTAGAGGTGAAGCACTTCGAACACTT 786
Db 241 SerGlyLysGlnLeuMetIleIleSerGluAspIleGlyGluAlaLeuAlaThrLeu 260
QY 787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCCTGTTTGT 846
Db 261 ValValAsnLysLeuArgGlyThrPheThrCysValGlyValLysAlaProGlyPheGly 280
QY 847 GAACCCCTCAAGCTATGCTGAAGATATGCTATCTTACTGAGGAGAGCAATATTT 906
Db 281 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyArgValValSer 300
QY 907 GAAGATCGTGTATTAAGCTTGAAATGTAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 966
Db 301 GluGluValGlyLeuLysLeuGluAsnAlaThrMetAspMetLeuGlyArgAlaArgGln 320
QY 967 GTAGTTATGCAAAAGAAATTAATCTATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1026
Db 321 ValArgIleSerLysGluGluThrIleValAspGlyAlaGlySerGlnAspAspIle 340
QY 1027 AAAGCTCCAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAAAGCTCAGATTATGCT 1086
Db 341 LysAlaArgIleAlaGlnIleArgGlnTyrGluGluSerThrSerGluPheAspLys 360
QY 1087 GAAAAAATTCAGAAACGCTTTCGCAAACTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1146
Db 361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleGlnValGly 380
QY 1147 GCTGCTACTGAACTGAATGAAGAGAGAGAGATCGTGTAGAGATGCTCTTAATGCA 1206
Db 381 AlaAlaThrGluThrGluLeuLysGluLysLysLeuArgIleGluAspAlaLeuAsnAla 400
QY 1207 ACAAGAGCTCCGTTGAAGAGGTATTCCTCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1266
Db 401 ThrArgAlaAlaValGluGluGlyIleValProGlyGlyIleThrAlaLeuValSerIle 420
QY 1267 ATTAAGTCTCTTGATGATTAATTAACCTGCTGATGATGATGATGATGATGATGATGATG 1326
Db 421 GlnLysAlaLeuAspGluValGluValProAlaGlyAspGluAlaThrGlyValAlaIle 440
QY 1327 ATCCGCTGCTCTTGAAGAGCTTTAGTCAANTTCGTCAAAATGCTGCAATGCTGCTATGAG 1386
Db 441 IleLysArgSerLeuGluGluProLeuArgGlnIleAlaAsnAlaAlaGlyTyrGluGly 460
QY 1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGTTTGGATTATTAATGCTGCATCA 1446
Db 461 SerValValValGluLysValLysAsnLeuProLysGlyGlnGlyPheAsnAlaAlaThr 480
```



```
QY 1447 GCAGAAATATGACAGCCTTATTAAAGCTGGTGTCTATTGATCTCTAAAGAAAGTTACACGATT 1506
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 GluValTyrGluAspMetIleAlaAlaGlyIleValAspProAlaLysValThrA9Ser 500
QY 1507 GCATTACAAATGACAGCATGAGTGGCTCTTACTTCTTAACCTACAGAAATGCGCTATTGCT 1566
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 AlaLeuGluAsnAlaAlaSerIleAlaAlaMetLeuLeuThrThrGluAlaIleValAla 520
QY 1567 GAAACACGAGCACTAAAGAGATATGCTATCCCTGGCGGTGATGGTGGTATGGGT 1626
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
521 AspLysProGlu---LysLysGluAlaProGlyMetProGlyMetProGlyMetGly 539
QY 1627 GGTATGGAC---GGTATG 1641
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 GlyMetAspMetGlyMet 545

RESULT 6
Q9AJB5 PRELIMINARY; PRT; 549 AA.
AC Q9AJB5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE groEL protein (60 kDa chaperonin) (protein Cpn60).
GN GROEL.
OS Pseudocalteromonas sp. PSIM3.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadales; Pseudalteromonas.
OX NCBI_TaxID=87791;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PSIM3;
RA Kuruu Y., Nakamura T.;
RT "Molecular chaperone of Psychrophile.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL; AB057417; BAB39465.1; -.
DR HSP; P06139; IGR.
DR InterPro; IPR001844; Chaperin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCPI.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 549 AA; 57346 MW; 7AFCC2E8A2F736AF CRC64;

Alignment Scores:
Pred. No.: 2,14e-86 Length: 549
Score: 1796.50 Matches: 360
Percent Similarity: 80.11% Conservative: 79
Best Local Similarity: 65.69% Mismatches: 108
Query Match: 60.53% Indels: 1
DB: 2 Gaps: 1

US-09-077-574A-1 (1-1647) x Q9AJB5 (1-549)
QY 1 ATGGCTTCTAAAGAAATCTTTTGTAGTGTAAAGCCGCTGAAACCTTTCAGAGGTGTA 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetAlaAlaLysGluValLeuPheAlaGlyAspAlaArgAlaLysMetLeuThrGlyVal 20
QY 61 GATAAACTGCAATGCTGTAAAGTAACTTCGACTGACCTAAGCCGCTAATGCTGTTATT 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGTCTTTTGGTTCCTCCAGTTATTACAAAGATGGTGTATCTGTGCAAAAGAAATT 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 AspLysSerPheGlySerProValIleThrLysAspGlyValSerValAlaLysGluIle 60
```

```
QY 181 GAACTTGAAGATAAGTTTCAAAATATATGGCGCTCAAAATGTTAAAGAGTAGCTCCAAA 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTGCTGGTGGTATGGAACATAACAAGCAACAGTCTCTTGCACAGCTATTAT 300
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 AlaAspAlaAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnSerIleVal 100
QY 301 CGTGAAGGTGTAACACTTGTAGCAGCTGGTGCCTAAATCCCTATGCCATTAAACGTCGCATA 360
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
QY 361 GATAAAGCTGTTGCTGCTTACTAAAGAACTAAGCGACATTACAAAGCCCTACTCTGTGAC 420
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 AspLysAlaIleIleAlaValAlaGluLeuLysAlaLeuSerValProCysAlaAsp 140
QY 421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAACCTCTCATACAACTAGGTAAAT 480
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspLysGluIleGlyAsp 160
QY 481 ATCATAGCTGAAGCTATGCTTAAGTTGGAAGGA---GGTCTTATCAGTTGAGGAA 537
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 IleIleAlaGluAlaMetGluLysValGlyArgAsnThrGlyValIleThrValGluGlu 180
QY 538 GCTAAAGGTCTTCAAACTACATTAGATGTGTTGAAGGAATGAAGTTTGACCGTGGCTAC 597
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 GlyGlnSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyr 200
QY 598 CTCTCTCCATCTTTGTAACATACTCTGAGAAATGGTTTGTGAACCTGATGATACCTTAT 657
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 LeuSerProTyrPheIleAsnAsnAlaGluLysGlyAlaValGluLeuAspAsnProPhe 220
QY 658 ATCTTTGTATCAGAAAAGATTACTAGCATGAACACATGCTACCAATCTTAGACAA 717
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 IleLeuLeuValAspLysLysValSerAsnIleArgGlnLeuLeuProThrLeuGluAla 240
QY 718 GTTGCTAAAGTAACCGTCCACTCTTATTATTGCTGAAGACGTAGAGGTGAAGCACTT 777
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 ValAlaLysAlaSerLysProLeuLeuIleAlaGluAspLeuGluGlyGluAlaLeu 260
QY 778 GCAACACTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAGAGCTCT 837
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 AlaThrLeuValValAsnAsnMetArgGlyIleValLysValSerAlaValLysAlaPro 280
QY 838 GCTTTGGTGAACCGCTTAAGCTATGCTTGAAGATATTCCTATCTCTTACTGAGAGGAA 897
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 GlyPheGlyAspArgGlyAlaMetLeuGlnAspIleAlaValLeuThrGlyGlyThr 300
QY 898 GCAATATTTCAGAGATCGTGTATATAAGCTTGAAGCTTGAAGCTTCTCTTTAGGAACA 957
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 ValIleSerGluGluIleGlyLeuGluLeuGluLysAlaThrValGluAspLeuGlyThr 320
QY 958 GCTAAACGTGTAGTTATTGACAAAGAAATATACTACTATCTGATGCTGTGGAAGATCA 1017
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
321 AlaLysArgValValIleThrLysAspAspThrThrIleIleAspGlyAlaGlyGluGlu 340
QY 1018 GAAGATATTAAAGCTCGAGTTAAACAATTCGTGCCACAAATTAAGAAACAAGCTCAGAT 1077
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 GluGlyIleThrGlyArgValSerGlnIleLysAlaGlnIleGluGluAlaThrSerAsp 360
QY 1078 TATGATCGTCAAAACTTCAAGAACGCTTTCGAAAACCTTGTGGTGGAGTAGCTGTATC 1137
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 TyrAspLysGluLysLeuGlnGluArgMetAlaLysLeuAlaGlyGlyValAlaValIle 380
QY 1138 CATGTTGGAGCTGCTACTGAAACTGAATGAAGAGAGAGAGGATCGCTAGAGATGCT 1197
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 LysValGlyAlaAlaThrGluIleGluMetLysLysLysAspArgValGluAspAla 400
QY 1198 CTAATCCAAACAGAGCTGCGGTGTAAGAGGATTTCTCCTGGTGGTGGTACTGCTTTT 1257
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 LeuHisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyValAlaLeu 420
```



```
1141 GTTGGAGCTGCTACTGAACTGAAATGAAAGAGAGAGGATCGTGTAGAGATGCTCTA 1200
Db ValGlyAlaThrGluValGluMetGlyGluLysLysAlaArgValGluAspAlaLeu 400
1201 AATGCAACAAGAGCTGGGTGGAAGAGGATTTGCTCCCTGGTGGTGTACTGCTTTTGTG 1260
Db HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuLe 420
1261 CGTCCATTAAAGCTTGTATGATGATTAACCTGCTGATGATGATGATGATGATGATGAT 1320
Db ArgValAlaSerLysIleAlaAspValGluValAlaAsnGluAspGlnLysHisGlyVal 440
1321 AATATCATCGCTGCTCTTCTTGAAGAGCTTTAGCTCAAAATGCTGCAAAATGCTGGCTAT 1380
Db ValIleAlaLeuArgAlaMetGluAlaProLeuArgGlnIleAlaThrAsnAlaGlyGlu 460
1381 GAAGGTCTATTGTTGAGAAAAGTTGCTGAAACCAAAAGATGTTTGGATTAAATGCT 1440
Db GluAlaSerValValAlaAsnAsnValLysAsnGlySerGlyAsnTyrGlyTyrAsnAla 480
1441 GCATCAGAGAAATGAGACCTTATTAAGCTGCTGATGATGATGATGATGATGATGATGAT 1500
Db GlyAsnAspThrTyrGlyAspMetLeuGluMetGlyIleLeuAspProThrLysValThr 500
1501 CGTATTGCATTACAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db ArgSerAlaLeuGlnPheAlaAlaSerIleAlaGlyLeuMetIleThrGluAlaMet 520
1561 ATTGCTGAAAACCCAGAACCTTAAAGATATGCTATGCTGCGGTGCTGATGCTGGTGGT 1620
Db ValAlaGluLeu-----ProLysAlaAspAlaPro---AspMetGlyGlyMetGlyGly 537
1621 ATGGCTGGTATGACGGTATG 1641
Db MetGlyGlyMetGlyGlyMet 544
```

RESULT 8

```
Q8GBB4 PRELIMINARY; PRT; 546 AA.
AC Q8GBB4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GROEL.
GN GROEL.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UMR;
RX MEDLINE=93323980; PubMed=7687328;
RA Ferat J.L., Michel F.;
RT "Group II self-splicing introns in bacteria.";
RL Nature 364:358-361(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UMR;
RA Ferat J.-L., Legouar M., Michel F.;
RT "An active group II intron has invaded the genus Azotobacter and is
RT inserted within the essential groEL gene.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057439; AAL25964.1; -.
SQ SEQUENCE 546 AA; 56875 MW; 8E1F4CC93E42A619 CRC64;
```

Alignment Scores:

Pred. No.:	1.47e-85	Length:	546
Score:	1780.50	Matches:	359
Percent Similarity:	79.63%	Conservative:	75
Best Local Similarity:	65.87%	Mismatches:	110
Query Match:	59.9%	Indels:	1

```
DB: 2 Gaps: 1
US-09-077-574a-1 (1-1647) x Q8GBB4 (1-546)
QY 1 ATGGCTTCTAAAGAAATCCCTTTTGTATGCTAAAGCCCGTGAAGAACTTTTCCAGAGGTGA 60
Db ValGlyAlaThrGluValGluMetGlyGlyGlyValValProGlyGlyGlyValValProGly 20
61 GATAAACTTCAATGCTGTTAAAGTAACTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 120
Db AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40
121 GAAAGTCTTTTGGTCTCCCGTATTTACAAAAGATGCTGCTATCTGTTGCAAAAGAAAT 180
Db AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60
181 GAACTTGAAGATAAGTTTGAATAATATGGCGCTCAATAGTGTAAAGAGTAGCTCCCAAA 240
Db GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 80
241 ACTAGCGATATTGCTGCTGATGGAACCTACACAGCAACAGCTCTTGCACAGCTATTAT 300
Db AlaAsnAspGluAlaGlyAspGlyThrThrAlaThrValLeuAlaGlnAlaIleVal 100
301 CGTGAAGCTGTAAACTTGTAGCAGCTGTCGTAATCTATGGCCATTAACAGTGGCATA 360
Db AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
361 GATAAAGCTCTTGTGCTGTACTTAAAGAACTTAAGCGACATTAACAGCTCTTCTCGTAC 420
Db AspLysAlaThrIleAlaIleValAlaGluLeuLysSerLeuAlaLysProCysSerAsp 140
421 CAAAAGAAATAGCTCAAGTTGGAACCAATTTCTGCAAACTCTGATACACAAATAGTAA 480
Db SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160
481 ATCATGCTGAAGCTATGCTGCTAAAGTTGGAAGAGGCTGTATTCACAGTTGAGGAAGCT 540
Db IleIleAlaGluAlaMetAsnLysValGlyLysGluGlyValIleThrValGluGluGly 180
541 AAAGGCTCTGAAACTACATAGATGTTGTTGAAGGAATGAAGTTTGACCGTGGCTACTCT 600
Db SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
601 TCTCCATCTTTGTAATACTCTCGTGAAGAAATGCTTGTGAACCTTGAATACCTTATATC 660
Db SerProTyrPheIleAsnLysProAspThrMetValAlaGluLeuAspAsnProLeuLeu 220
661 CTTTGTAAATGAGAAAAGATTTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAAT 720
Db LeuLeuValAspLysLysIleSerAsnIleArgGluLeuLeuProValLeuGluAlaVal 240
721 GCTAAAGTAAACCGTCCACTCTCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGA 780
Db AlaLysSerGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260
781 ACATTGTAGTCAATAGCTCCCTGAGCACTCCCAAGTTCTAGCCGTAAGCTCTCTGGT 840
Db ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280
841 TTTGTTGAACCGCTAAAGCTATGCTTGAAGATATTGCTATCTCTTACTCTGAGGAGAGCA 900
Db PheAsnAspArgGlyLysAlaMetLeuGlnValIleAlaIleLeuThrGlyAlaThrVal 300
901 ATATTTCAAGATCGTGTATTAAGCTTGAAGATTTGAAGCTTGTCTTCTTTAGGAACAGCT 960
Db IleSerGluLysValGlyLeuSerLeuGluSerAlaThrLeuGluHisLeuLysPro 320
961 AAAGCTGTATTATGACAAAGAAATACTACTATCTGTTGATCGTGTGCTGAAATCAGAA 1020
Db LysAlaLeuValLeuAsnLysGluAsnThrThrIleMetHisGlyAlaGlnAla 340
1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCAAAATTCGTGCAAAACACAGCTCAGATTAT 1080
```


Db 281 AspArgGlySalMetLeuGluAspIleAlaIleLeuThrGlyGluValIleSer 300
 QY 907 GAAGATCGTGTATATAAGCTTGAATGTAGCTTCTTTTAGGACAGCTAAACGT 966
 Db 301 GluGluLeuGlyArgGluLeuLysSerAlaThrIleAlaSerLeuGlyArgAlaSerLys 320
 QY 967 GTAGTTATTGACAAAGAAATCTACTATCTGTTGATGGTGGTGAATAATCAGAAGATATT 1026
 Db 321 ValValValThrLysGluAsnThrThrIleValAspGlyAlaGlyAspSerGluArgIle 340
 QY 1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTTGAAGAAACAGCTCAGATTATGCTT 1086
 Db 341 LysAlaArgIleAsnGlnIleArgAlaGlnLeuGluGluThrThrSerGluPheAspArg 360
 QY 1087 GAAAACTTCAGAACGCTCTTGCAGAACTTGTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
 Db 361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValArgValIleLysValGly 380
 QY 1147 GCTGCTACTGAACTGAAATGAAAGAGAGAGATCGTGTAGAGATGCTCTAAATGCA 1206
 Db 381 AlaAlaThrGluThrGluLeuLysGluArgLysLeuArgIleGluAspAlaLeuAsnSer 400
 QY 1207 ACAAGAGCTCGGTTGAAGAGGATTTGTCCCTGGTGGTGGTACTGCTTTGTCCTCC 1266
 Db 401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyThrAlaLeuMetAsnVal 420
 QY 1267 ATTAAGTCTTGATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGATGAT 1326
 Db 421 TyrAsnLysValAlaAlaIleGlu--AlaGluGlyAspGluAlaThrGlyValLysIle 439
 QY 1327 ATCCGCTGCTTCTTCAAGAGCTTTACGTCAAATTCGTCCTCAATGCTGCTGCTGCTGCT 1386
 Db 440 ValLeuArgAlaIleGluProValArgGlnIleAlaGlnAsnAlaGlyLeuGluGly 459
 QY 1387 TCTATTGTTGTGAAAAGTTCGTGACCAACAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
 Db 460 SerValIleValGluArgLeuLysThrGluLysProGlyIleGlyPheAsnAlaAlaThr 479
 QY 1447 GGAGATATGAGACCTTATAAGCTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 1506
 Db 480 GlyGluTrpValAspMetIleGluAlaGlyIleValAspProThrLysValThrArgSer 499
 QY 1507 GCATTACAAATCAGCATCAGTACTCTTACTTCTTACTTCTTACTTCTTACTTCTTACTTCT 1566
 Db 500 AlaLeuGlnAsnAlaAlaSerValAlaAlaMetPheLeuThrThrGluAlaValAla 519
 QY 1567 GAAAAACCAAGACCTTAAAGATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1626
 Db 520 AspLysProGluGluAsnLys-----GlyGlyAsnAlaGlyMetPro 533
 QY 1627 GGTATGACGGTATG 1641
 Db 534 AspMetGlyGlyMet 538

RESULT 10

O31198 PRELIMINARY; PRT; 546 AA.
 AC O31198;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Heat shock protein 58 (60 kDa chaperonin) (Protein Cpn60) (groEL protein).
 GN HSP58.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 OX NCBI_TaxID=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HY-1;
 RA Kim M.J., Ahn B.Y.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND

CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 DR EMBL; AF032910; AAB86965.1; -.
 DR HSP; P06139; 1GRL.
 DR InterPro; IPR001844; Chaperonin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR Pfam; PF00118; cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONING6.
 DR PRINTS; PR00304; TCOMPLEXTCPL.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 KW ATP-binding; Chaperone.
 SQ SEQUENCE 546 AA; 57969 MW; 15BC9EC0A8E2A6E3 CRC64;

Alignment Scores:

Pred. No.: 6,63e-85 Length: 546
 Score: 1768.00 Matches: 352
 Percent Similarity: 80.07% Conservative: 86
 Best Local Similarity: 64.35% Mismatches: 101
 Query Match: 59.57% Indels: 8
 DB: 2 Gaps: 3

US-09-077-574A-1 (1-1647) x O31198 (1-546)

QY 7 TCTAAGAAATCTTTTGTGATGCTAAAGCCGCTGAAAGAACTTTCACGAGGTGATATAA 66
 Db 2 AlaLysAspIleGluTrpAsnGluThrAlaArgLysLeuLeuGluGlyValAsnLys 21
 QY 67 CTTGCAAAATGCTGTTAAAGTAACACTTGGACCTTAAAGCGCGTAAATGCTTATTGAAAAG 126
 Db 22 LeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIleAspLys 41
 QY 127 TCTTTTGGTTCCTCCAGTTTATACAAAGATGGTGTATCTGTTCCAAAGAAATGTAACCTT 186
 Db 42 LysPheGlyAlaProThrIleThrLysAspGlyValThrValAlaLysGluIleGluLeu 61
 QY 187 GAAGATAAGTTTGAAATATGCGCGCTCAATGTTTAAAGAGTAGTCTCCAAACTAGC 246
 Db 62 GluAspProLeuGluAsnMetGlyAlaGlnMetValLysGluValSerThrLysThrAsn 81
 QY 247 GATATTGCTGCTGATCGAACTACACAGCAAGCTCTTGCAACAGCTATTTATCTGTAA 306
 Db 82 AspValAlaGlyAspGlyThrThrAlaThrIleLeuAlaGlnSerIleIleAsnGlu 101
 QY 307 GGTGTAACACTTGTAGCAGCTGCTGATATCTTATGCTATGCTTAAACCTGATAGATAAA 366
 Db 102 GlyLysAsnValThrAlaGlyAlaAsnProMetSerLeuLysArgGlyIleAspLys 121
 QY 367 GCTGTTGTTGCTGTTACTTAAAGAACTAAGGACATTACAAAGCTACTCGTGACCAAAA 426
 Db 122 AlaValThrAlaAlaValGluSerIleGlnLysArgAlaValLysIleGluAsnLysLys 141
 QY 427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACACAACTAGTATATATCAT 486
 Db 142 AspIleAlaAsnValAlaSerIleSerAlaAsnAsnAspAsnThrIleGlyAsnLeuIle 161
 QY 487 GCTGAAAGCTATGCTTAAAGTTGAAAAGGAGGTGTTTATCACAGTTGAGGAAGCTAAAGGT 546
 Db 162 AlaAspAlaMetAspLysValGlyLysAspGlyValIleThrValGluGluAlaLysSer 181
 QY 547 CTTGAAACTACATTAGATGTTGGAAGAAATGAAGTTCACCGTGCCTACCTCTCTCCA 606
 Db 182 IleGluThrThrLeuAspValValGluGlyMetGlnPheAspArgGlyTrpIleSerPro 201
 QY 607 TACTTTGTAACCTAATCTCGAGAAATGTTTGGAACCTTGATAACCTTATATCTTTTGT 666
 Db 202 TyrMetValThrAspAlaGluSerMetValAlaThrLeuAsnAspPropheIleLeuIle 221
 QY 667 AATGAGAAAGATTTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAAGTTGTCTAA 726
 Db 222 TyrAspLysLysIleSerMetLysAspLeuIleHisIleLeuGluLysValAlaGln 241


```
QY 547 CTTGAAACTACATTAGATGGTGTGAAGGAATGAAGTTTGACCGTGCGTACCTCTCTCCA 606
Db 181 PheThrThrGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrAlaSerPro 200
QY 607 TACTTTGTAATACTCCTGAGAAATGGTTGTGAACCTGATACACCTTATATCTTTTGT 666
Db 201 TyrMetIleThrAspThrGluLysMetGluAlaValLeuGluAsnProTyrIleLeuIle 220
QY 667 AATGAGAAAAGATTACTAGCATGAAGACATGCTACCAATCTTAGAACCAAGTTGCTAAA 726
Db 221 ThrAspLysLysIleSerAsnIleGlnAspLeuLeuProIleLeuGluGlnValValGln 240
QY 727 GTAAACCGTCCACTCTCTATTATTGCTGAAGAGCTAGAGGTGAAGCACTTGCAACACTT 786
Db 241 GlnGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 260
QY 787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAGCTCTGTTTGTGT 846
Db 261 ValValAsnLysLeuArgGlyThrPheThrAlaValAlaValLysAlaProGlyPheGly 280
QY 847 GAACGCGTTAAAGCTATGCTTGAAGATATGCTTACTGAGGAGCAATATTT 906
Db 281 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluValIleSer 300
QY 907 GAAGATCGTGATATAAGCTTGAAGCTTGAAGTGTGCTTCTTTAGCAACAGCTTAAACGT 966
Db 301 GluGluLeuGlyArgGluLeuLysSerThrThrIleAlaSerLeuGlyArgAlaSerLys 320
QY 967 GTAGTTATTGACAAAGAAATATCTATCTGTTGATGCTGCTGAAATCAGAGATATT 1026
Db 321 ValValValThrLysGluAsnThrThrIleValGluGlyAlaGlyAspSerAspArgIle 340
QY 1027 AAGCTCGATTAAACAATCTGTCACAAATTTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db 341 LysAlaArgIleAsnGlnIleArgAlaGlnLeuGluGluThrThrSerGluPheAspArg 360
QY 1087 GAAAACTTCAAGAACTCTTGAAGAACTTGTGGTGGAGTAGCTGTATCCATGTTGGA 1146
Db 361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyValAlaValIleLysValGly 380
QY 1147 GCTGCTACTGAAACTGAAATGAAGAGAGAGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db 381 AlaAlaThrGluThrGluLeuLysGluArgLysLeuArgIleGluAspAlaLeuAspSer 400
QY 1207 ACAAGAGCTCGGTTGAAGAGTATTGCTCCGTGGTGGTGGTGGTGGTGGTGGTGGTCC 1266
Db 401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyThrAlaLeuMetAsnVal 420
QY 1267 ATTAAGTCTCTGATGATATTAAACCTGCTGATGATGATGATGATGATGATGATGATC 1326
Db 421 TyrSerLysValAlaAlaIleGlu--AlaGluGlyAspGluAlaThrGlyValLysIle 439
QY 1327 ATCCGCTGCTCTCTGAAAGAGCCTTTACGTCCTCAATGCTGCTGCTGCTGCTGCTGCTG 1386
Db 440 ValLeuArgAlaIleGluGluProValArgGlnIleAlaGlnAsnAlaGlyLeuGluGly 459
QY 1387 TCATTGTTGTGAAAAGTCTGTGAACCAAGATGGTGGTGGTGGTGGTGGTGGTGGTGGTCA 1446
Db 460 SerValIleValGluArgLeuLysThrGluLysProGlyIleGlyPheAsnAlaAlaThr 479
QY 1447 GGAGATATGAGACCTTATTAAAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1506
Db 480 GlyGluTrpValAsnMetIleGluAlaGlyIleValAspProThrLysValThrArgSer 499
QY 1507 GCATTACAAAATGAGCATCAGTACGCTCTCTTACTTCTAACTACAGAAATGCGTATTGCT 1566
Db 500 AlaLeuGlnAsnAlaAlaSerValAlaAlaMetPheLeuThrThrGluAlaValValAla 519
QY 1567 GAAAAACCAAGCACTAAAAAAGATATGCTATGCTGGCGGTGGTGGTGGTGGTGGTGGTGG 1626
Db 520 AspLysProGluGluAsnLys-----GlyGlyAsnAsnGlyMetPro 533
```

```
QY 1627 GGATGACGGTATG 1641
Db 534 AspMetGlyGlyMet 538
RESULT 12
Q93F08 PRELIMINARY; PRT; 545 AA.
ID Q93F08
AC Q93F08;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE groEL-like protein (60 kDa chaperonin) (Protein Cpn60) (groEL
DE protein).
DE GROEL.
GN Enterobacter aerogenes (Aerobacter aerogenes).
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=548;
RN [1]
SEQUENCE FROM N.A.
RA Yoshida N., Oeda K., Watanabe E., Mikami T., Fukita Y., Nishimura K.,
RA Komai K., Matsuda K.;
RT "A GroEL homolog produced by endosymbionts in antlions as a paralytic
RT toxin against insects.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -I- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF306521; AAL09389.1; -
DR InterPro: IPR001844; Chaperin Cpn60.
DR InterPro: IPR002423; Cpn60/TCPI-1.
DR Pfam: PF001118; cpn60_TCPI.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 545 AA; 57023 MW; DB3B7A017DEDA85F CRC64;
Alignment Scores:
Pred. No.: 1-52e-83 Length: 545
Score: 1742.00 Matches: 353
Percent Similarity: 78.02% Conservative: 73
Best Local Similarity: 64.65% Mismatches: 116
Query Match: 58.69% Indels: 4
DB: 2 Gaps: 2
US-09-077-574A-1 (1-1647) x Q93F08 (1-545)
QY 1 ATGGCTTCTTAAGAAATCCTTTTGTATGCTAAAGCCGCGTAAAGAACTTTCCAGCGGTGA 60
Db 1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20
QY 61 GATAAAGCTGCAATGCTGTTAAAGTAACACTGGACCTAAAGCCGCTAATGCTGTTATT 120
Db 21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
QY 121 GAAAGTCTTTTGGTCTCCCGAGTTATTACAAAAGATGCTGCTATCTGTTGCAAAAGAAATT 180
Db 41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60
QY 181 GAACCTTGAGATTAAGTTTGAAATATGGCGCTCAAAATGTTAAAGTAGTAGTCTCCCAA 240
Db 61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80
QY 241 ACTAGCGATATTCTGCTGATGCAACTACACAGACAGACTCCTTCACAAGCTATTAT 300
Db 81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100
QY 301 CGTGAAGGTGTAACACTTGTAGCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
```

101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120
361 GATTAAGCTGTGTGCTTACTAAAGAACTAAGCAGCATTTACAAAGCCCTACTCGTGAC 420
121 AspLysAlaValAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140
421 CAAAAGAAATAGCTCAAGTGTGAACCATTTCTGCAAACTCTGTATACCAACATAGGTAAT 480
141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160
481 ATCATAGCTGAAGCTATCGCTAAAGTGTGAAGAGAGGCTGTATACAGTTGAGGAAGCT 540
161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180
541 AAAGCTCTTGAACACTACATAGATGTGTGAGGAAGTGAAGTTGACGGTGGCTACCTC 600
181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
601 TCTCATATCTTTGTAACCTAATCTCAGAAAAATGGTTGTGAACCTTGATAACCTTATATC 660
201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220
661 CTTTGTATAGAAAGAAATGACTAGCATGAAAGACATGCTACCAATCTTAGAACAGTT 720
221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240
721 GCTAAAGTAAACCGTCCACTCTTATTATTGCTGAAGAGCTAGAGGTGAAGCACTTGA 780
241 AlaLysAlaGlyLysProLeuLeuIleAlaGluAspValGluGlyValGluAlaLeuAla 260
781 ACACCTGTAGTCAATTAAGCTCGCTGAGCAGCTCCCAAGTTGTAGCCGTAAAGCTCTCGT 840
261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280
841 TTTGTGTAACCGCTGAAGCTATGCTGAAGATATGCTATCTTACTTACGAGGAGCA 900
281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300
901 ATATTGTAAGCTCGGTGATAAGCTTGAAGTGAAGTGAAGTCTCTCTTCTAGGACACT 960
301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320
961 AAACGTGTAGTTATTGACAAAGAAATACTACTATCTGTGTGTGCTGCGAAAAATCAGAA 1020
321 LysArgValValIleAsnLysAspThrThrThrIleAspGlyValGlyAspGluAla 340
1021 GATATTAAGCTCGAGTAAACAAATTCGTGCACAAATGGAAGAACAGCTCAGATTAT 1080
341 AlaIleGlnGlyArgValThrGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360
1081 GATCGTGAAGAACTTCAAGAGCTCTGCAAACTGTGTGTGAGTAGCTGTATCCAT 1140
361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyValAlaValIleLys 380
1141 GTTGAGCTGCTPACTCAAACTCAAAATGAAAGAGAAAGAGTGTGTGAGAAATGCTCTTA 1200
381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400
1201 AATGCAACAGAGCTCGGTTGAAGAGTATTGTCCTCGGTGCTGCTGCTGCTTTC 1260
401 HisAlaThrArgAlaAlaValGluGluValAlaAlaGlyValGlyValAlaLeuIle 420
1261 CGCTCCATTAAAGCTCTGATGATATTAAACCTGCTGATGATGATGATGCTGCTGCT 1320
421 ArgValAlaSerLysIleAlaGlyLeuLysGlyGlnAsnGluAspGlnAsnValGlyIle 440
1321 AATATCATCGCTGCTTCTTGAAGAGCTTTTACGTCAAAATGCTGCAAAATCGCTGCTAT 1380
441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460
1381 GAAGGTTCTATTGTTGTAAGAAAGCTCTGTGAACCAAGATGCTTTGGATTTATGCT 1440
461 GluProSerValAlaAlaAsnThrValLysAlaGlyAspGlyAsnTyrGlyTyrAsnAla 480

1441 GCATCAGCAGAAATATGAAGACCTTATAAGCTGTGTGCTTATGATCTCTAAAAAGTTACA 1500
481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500
1501 CGTATTGCTATTACAAATGCAGCATCAGTACCTCTTACTTCTTAACACAGAAATGCGCT 1560
501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520
1561 ATTGCTGAAACACCAACCTTAAAAAGATATGCTATGCTGCGC-----GCTGGTATG 1614
521 ValThrAspLeu-----ProLysSerAspAlaProAspLeuGlyAlaAlaGlyMet 538
1615 GGTGCTATGGTGTATG 1632
539 GlyGlyMetGlyGlyMet 544
RESULT 13
O33688 PRELIMINARY; PRT; 538 AA.
AC O33688;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE GROEL protein (60 kDa chaperonin) (Protein Cpn60) (Fragment).
GN GROEL.
OS primary endosymbiont of *Sitophilus oryzae*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae.
OX NCBI_TaxID=2342;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-SF:
RX MEDLINE=98042468; PubMed=9367844;
RA Charles H., Heddi A., Guillaud J., Nardon C., Nardon P.;
RT "A molecular aspect of symbiotic interactions between the weevil
RT *Sitophilus oryzae* and its endosymbiotic bacteria: over-expression of a
RT chaperonin."
RL Biochem. Biophys. Res. Commun. 239:769-774(1997).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC HSP; AF005236; AAB97670.1; --.
DR HSP; P06139; IJON.
DR InterPro; IPR001844; Chaperin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONING.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
FT NON TER 538 538
SQ SEQUENCE 538 AA; 56683 MW; EFD17668C0F09D15 CRC64;

Alignment Scores:

Aligned No.: 3, 128-83 Length: 538
Score: 1736.00 Matches: 348
Percent Similarity: 78.86% Conservative: 81
Best Local Similarity: 63.97% Mismatches: 109
Query Match: 58.45% Indels: 6
DB: 2 Gaps: 2

US-09-077-574A-1 (1-1647) x O33688 (1-538)

OY 1 ATGGCTTCTAAGAAATCTTTTTCATGCTAAAGCCGCTGAAAACTTTTCAGGAGTGA 60
DB 1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20
OY 61 GATAAATCTGCAATGCTGTGTAAAGTAAACACTTGACCTTAAAGCGCGTAAATGCTTAT 120
DB 61 GATAAATCTGCAATGCTGTGTAAAGTAAACACTTGACCTTAAAGCGCGTAAATGCTTAT 120

us-09-077-574a-1.rspt

Mon Feb 2 12:04:08 2004

Search completed: January 28, 2004, 13:28:47
Job time : 98 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:10:00 ; Search time 47 Seconds
(without alignments)
1850.684 Million cell updates/sec

Title: US-09-077-574A-2

Perfect score: 2708

Sequence: 1 MASKEILLFDKAREKLSRGV.....KDMFPGGGMGGMGMDGMY 548

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2708	100.0	548	AAW16678	Lawsonia intracell
2	1864.5	68.9	545	20 AAY23930	Consensus mino aci
3	1848	68.2	549	20 AAY23909	Amino acid sequenc
4	1825.5	67.4	547	22 AAB69060	Pseudomonas aerugi
5	1820.5	67.2	547	16 AAR67381	L. pneumophila Htp
6	1819	67.2	574	22 AAB59061	ptfChisB expressio
7	1814.5	67.0	545	20 AAY23915	Amino acid sequenc
8	1812.5	66.9	544	21 AAY75747	Neisseria meningit
9	1805.5	66.7	544	24 ABP77245	N. gonorrhoeae ami

10	1769.5	65.3	544	23	AAE22070	Francisella tulare
11	1750.5	64.6	544	21	RAY75745	Neisseria gonorrhe
12	1749	64.6	545	23	ABB99014	Detritmental organi
13	1746	64.5	545	21	AAB19080	Amino acid sequenc
14	1746	64.5	548	19	AAW65070	E. coli GroEL prot
15	1746	64.5	548	20	AAW95001	W0902989 Seg ID 1
16	1746	64.5	548	22	AAB50536	Escherichia coli G
17	1746	64.5	548	23	ABB99010	Detritmental organi
18	1744	64.4	545	23	ABB99013	Detritmental organi
19	1743.5	64.4	544	20	AAZ21905	Amino acid sequenc
20	1742	64.3	545	23	ABB99009	Detritmental organi
21	1742	64.3	545	23	ABB99012	Detritmental organi
22	1742	64.3	545	23	ABB99016	Detritmental organi
23	1741	64.3	548	19	AAW61378	GroEL N-terminal r
24	1741	64.3	548	21	AAZ99621	Escherichia coli g
25	1739	64.2	545	23	ABB99015	Detritmental organi
26	1738	64.2	545	23	ABB99011	Detritmental organi
27	1736.5	64.1	548	20	AAZ23908	Amino acid sequenc
28	1735	64.1	548	16	AAR67382	E. coli GroEL gene
29	1734	64.0	545	23	ABB99017	Detritmental organi
30	1733	64.0	545	20	AAZ23901	Streptococcus pneu
31	1732	64.0	542	23	ABB49241	Listeria monocytog
32	1729.5	63.9	544	20	AAZ23903	Streptococcus pyog
33	1723.5	63.6	546	14	AAR41661	H. pylori hsp60.
34	1723.5	63.6	546	14	AAR41200	Hsp. Helicobacter
35	1723	63.6	544	12	AAZ13337	HyPB protein. Chl
36	1720.5	63.5	546	19	AAW98776	H. pylori GHPO 118
37	1720	63.5	548	23	ABB99019	Detritmental organi
38	1720	63.5	548	23	ABB99021	Detritmental organi
39	1719	63.5	548	23	ABB99018	Detritmental organi
40	1718	63.4	544	16	AAR67383	C. psittaci HyPB g
41	1717	63.4	548	23	ABB99020	Detritmental organi
42	1716.5	63.4	547	23	AAU76192	Human PI protein (
43	1716.5	63.4	547	23	ABG74586	E. coli groEL prot
44	1710	63.1	544	22	AAE11757	Chlamydia pneumonia
45	1710	63.1	544	23	ABB94272	Chlamydia pneumonia

ALIGNMENTS

RESULT 1

AAW16678
ID AAW16678 standard; Protein; 548 AA.

XX AAW16678;

AC AAW16678;

XX 20-AUG-1997 (first entry)

DT Lawsonia intracellularis GroEL.

DE Intestinal disease; porcine proliferative enteropathy; vaccine;

KW GroEL; immunotherapy; antibody; diagnosis; heat shock protein.

OS Lawsonia intracellularis.

XX WO9720050-A1.

XX 05-JUN-1997.

XX 29-NOV-1996; 96WO-AU00767.

XX 30-NOV-1995; 95AU-0006911.

XX 30-NOV-1995; 95AU-0006910.

XX (DARA-) DARATECH PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

XX Hasse D, Panaccio M;

XX WPI; 1997-310605/28.

XX N-PSDB; AAT69201.

PT Vaccine for treating or preventing Lawsonia intracellularis
 PT infection especially in pigs, containing non-pathogenic form of
 PT bacterium or its components
 XX
 XX
 PS Claim 10; Page 38-42; 94pp; English.
 XX
 CC The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis
 CC can be used in vaccines to protect birds and animals against
 CC intestinal diseases, esp. to protect pigs against porcine
 CC proliferative enteropathy. It is the expression product of a DNA
 CC molecule (AAW69201) obtd. by screening an L. intracellularis library
 CC with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)
 CC and other L. intracellularis polypeptides (AAW16680-85) can be
 CC administered as recombinant polypeptides or expressed as
 CC recombinant vaccines utilising bacterial, fungal or viral vectors.
 CC Antibodies raised against the polypeptides may be useful in
 CC immunotherapy, diagnosis of infection and detection.
 XX
 SQ Sequence 548 AA;

Query Match 100.0%; Score 2708; DB 18; Length 548;
 Best Local Similarity 100.0%; Pred. No. 3.4e-170;
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGKGRNVIEKSGFSPVITKDGVSVAKEI 60
 Db 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGKGRNVIEKSGFSPVITKDGVSVAKEI 60
 QY 61 ELEDKFNMGAMQVKEVAPKTSIADGGTTATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
 Db 61 ELEDKFNMGAMQVKEVAPKTSIADGGTTATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
 QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEA 180
 Db 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEA 180
 QY 181 KGLTLLDVVEGKMFDRGYLSPFVTNPEKMWCELDNPNYLCNEKTTSMKMLPILEQV 240
 Db 181 KGLTLLDVVEGKMFDRGYLSPFVTNPEKMWCELDNPNYLCNEKTTSMKMLPILEQV 240
 QY 241 AKVNRPLIIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGERRKAMLEDIAITLGG 300
 Db 241 AKVNRPLIIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGERRKAMLEDIAITLGG 300
 QY 301 IPEDRGIKLENVSLSSIGTAKRVVIDKENTTIIVDGAKSEDIKARVKQIRAOIEETSSDY 360
 Db 301 IPEDRGIKLENVSLSSIGTAKRVVIDKENTTIIVDGAKSEDIKARVKQIRAOIEETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGATETEMKEKKDORVEDALNATRAAVEEGIVPGGGTAFV 420
 Db 361 DREKLOERLAKLVGGVAVIHVGATETEMKEKKDORVEDALNATRAAVEEGIVPGGGTAFV 420
 QY 421 RSIKVLDDIKPADDDDELAGLNIIRRSLEELRQIAANAGYEGSVIWEKVEPKDGFQFNA 480
 Db 421 RSIKVLDDIKPADDDDELAGLNIIRRSLEELRQIAANAGYEGSVIWEKVEPKDGFQFNA 480
 QY 481 ASGEYEDLIKAGVIDPKKVTIRIALONAAVASLILLTTECAIAEKPEPKDMPGPGMG 540
 Db 481 ASGEYEDLIKAGVIDPKKVTIRIALONAAVASLILLTTECAIAEKPEPKDMPGPGMG 540
 QY 541 MGGMDGMY 548
 Db 541 MGGMDGMY 548

RESULT 2
 ID AAY23930
 XX AAY23930 standard; Protein; 545 AA.
 AC AAY23930;
 XX
 XX 22-SEP-1999 (first entry)
 DT
 XX

DE Consensus amino acid sequence of a heat shock protein.
 XX Heat shock protein; Hsp; immune response; immunological carrier;
 KW cancer control; tumour; sarcoma; cancer; gene therapy.
 XX
 OS Synthetic.
 XX
 XX WO9935270-A1.
 PN 15-JUL-1999.
 PD 29-DEC-1998; 98WO-CA01203.
 PF 31-DEC-1997; 97US-0001737.
 PR
 XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 PA Mizzen L, Wisniewski J;
 XX MPI; 1999-430397/36.
 DR
 XX New nucleic acid encoding heat shock protein-60 from Streptococcus,
 PT useful in vaccines, as carriers for other immunogens, as anticancer
 PT agents and for diagnosis
 XX
 PS Disclosure; Fig 10A-E; 176pp; English.
 XX
 CC AAY23905-30 represent heat shock proteins (Hsps). The specification
 CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
 CC fragments, variants and fusion proteins, are used to elicit or enhance
 CC an immune response against Streptococcus, and to elicit a similar
 CC response to a target antigen fused to the protein. Unlike other
 CC immunological carriers, Hsp60 proteins are not immunosuppressive so
 CC provide an increased response to any conjugated or fused antigen. Also,
 CC where used for cancer control, they lack the side effects associated
 CC with endotoxins. They can also be used to detect specific antibodies
 CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
 CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
 CC polynucleotide is used for recombinant production of the protein, as
 CC a source of primers and probes for detecting streptococci in standard
 CC hybridization/amplification assays, and therapeutically in gene
 CC therapy vectors.
 XX
 SQ Sequence 545 AA;

Query Match 68.9%; Score 1864.5; DB 20; Length 545;
 Best Local Similarity 69.8%; Pred. No. 1.1e-114;
 Matches 380; Conservative 74; Mismatches 85; Indels 7; Gaps 5;
 QY 3 SKELIFDAKAREKLSRGVDKLANAVKVTGLGKGRNVIEKSGFSPVITKDGVSVAKEI 62
 Db 2 AKDIKTGEAEARRAMLRGVNALADAVKVTGLGKGRNVIEKSGFAPITTKDGVTVAKEI 61
 QY 63 EDRFENMGAMQVKEVAPKTSIADGGTTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 122
 Db 62 EDRFENMGAMQVKEVAPKTSIADGGTTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 121
 QY 123 AVAVVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEAKG 182
 Db 122 AVDAVVEELKAIKAPVETKEIAQVATISANGDEEIGELIAEMAEKVGKGVITVEEGKT 181
 QY 183 LETTLDVVEGKMFDRGYLSPFVTNPEKMWCELDNPNYLCNEKTTSMKMLPILEQVAK 242
 Db 182 LETELEVVEGQFDRGYISPFYITDSEKQAELEDPILLLTDDKISINIDLLPVEEVAQ 241
 QY 243 VNRPLIIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGERRKAMLEDIAITLGGGAIF 302
 Db 242 AGKPLIIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGERRKAMLEDIAITLGGQVIS 301
 QY 303 EDRIKLENVSLSSIGTAKRVVIDKENTTIIVDGAKSEDIKARVKQIRAOIEETSSDYDR 362
 Db 302 EELGSLSEDATLEDLGQAKKVVVTKDOTTIIVDGAGDAA-TAGRVAQIRSQIEESTSDYDK 360

363 EKQERLAKLVGVAVIHVGAATETEMKEDKORVEDALNATRAAAVEGIVPGGTAFFVRS 422
361 EKQERLAKLVGVAVIKVGATEVELKERDORVEDALNATRAAAVEGIVPGGVALIRA 420
423 IKVLDLKPADDLAGLNIIRSLRQIAANAGYEGSVVVEKVPKDGFGFNAAS 482
421 APALDKLKTENGDEATGNIIVLRALEAPLQIAENAGLEGSVVVEKVKNSAG-GYNAAT 479
483 GEYEDLIKAGVIDPKVTRIALQNAASVASILLITTECAIAEKPEPKDMP--MPC--GGM 538
480 GEYVDMIAAGIIDPTKVRSLQNAASVASLMLTTEAVVVDKPE-KEAAPAGMPGMMGM 538
539 GGMGM 544
539 GGMGM 544

RESULT 3
ID AAY23909 standard; Protein; 549 AA.
AC AAY23909;
XX
XX 22-SEP-1999 (first entry)
XX
XX Amino acid sequence of a heat shock protein.
XX
XX Heat shock protein; Hsp; immune response; immunological carrier;
KW cancer control; tumour; sarcoma; cancer; gene therapy.
XX
XX Legionella pneumophila.
XX
XX WO9935270-A1.
XX
XX 15-JUL-1999.
XX
XX 29-DEC-1998; 98WO-CA01203.
XX
XX 31-DEC-1997; 97US-0001737.
XX
XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
XX Mizzen L, Wisniewski J;
XX
XX WPI; 1999-430397/36.
XX
XX New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT useful in vaccines, as carriers for other immunogens, as anticancer
PT agents and for diagnosis
XX
XX Disclosure; Fig 10A-E; 176pp; English.

XX
XX AAY23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.

XX Sequence 549 AA;

Query Match 68.2%; Score 1848; DB 20; Length 549;
Best Local Similarity 67.3%; Pred. No. 1.4e-113;
Matches 367; Conservative 80; Mismatches 96; Indels 2; Gaps 1;

QY 3 SKELFDKAREKLSRGVDKLANAVKVTLPKGRNVLEKSGSPVITKDGVSVAKEIETL 62
DB 2 AKELRFGDARLQWLAGVNLADAVQVTMGPRGNVLEKSYGAPTIVTKDGVSVAKEIETL 61
QY 63 EDKFNMGAMQVKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAAGNPMIAIKRGIDK 122
DB 62 EHRFMNGAMQVKEVASKTSDTAGDGTTTATVLAARSILVEGHKA VAAAGNPMIAIKRGIDK 121
QY 123 AVAVTKELSDITKPTDRDQKIAQVGTISANSDDTTIGNIIAENAKVKGKGVITVEEAKG 182
DB 122 AVLAVTKLQAMSKPCKDSKAIQVGTISANSDEAIGAIIAEAMEKVGKGVITVEDGNG 181
QY 183 LETTLDVVVEGKFDGRGYSVFTNPEKMWCELDNPYILCNEKKTSMKMLPILLEQVAK 242
DB 182 LENELSVVEGQFDRGYISPIFINNQNMSELEHPFILLVKKVSSIREMLSVLEGVAK 241
QY 243 VNRELLIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
DB 242 SGRPELLIIAEDVEGEALATLVNMMRGIVKCAVKAPGFGERRKAMQLDIAILTKGVIS 301
QY 303 EDRGIKLENVLSLGTAKRVVIDKENTTIIVDGAKSSEDIKARVKQIPRAQIEETSSDYDR 362
DB 302 BEIGKSLGATLEDLGSARKIVVTKENTTIIDGEGKATEINARITQIRAOEETTSYDR 361
QY 363 EKQERLAKLVGVAVIHVGAATEMKEKKDRVEDALNATRAAAVEGIVPGGTAFFVRS 422
DB 362 EKQERLAKLVGVAVIKVGATEVELKERDORVEDALNATRAAAVEGIVPGGVALIRA 421
QY 423 IKVLDLKPADDLAGLNIIRSLRQIAANAGYEGSVVVEKVPKDGFGFNAAS 482
DB 422 QKALDSLKGDNDQNMGINILRRAIESPMRQIVTNAGYEAASVNVKVAEHKDNVGFNAAT 481
QY 483 GEYEDLIKAGVIDPKVTRIALQNAASVASILLITTECAIAEKPEPKDMPGMMGM 542
DB 482 GEYVDMVEMGILDPTKVRSLQNAASVASLMLTTECMVADL--PKKEEGVAGDMGGM 539
QY 543 GMDGM 547
DB 540 GMDGM 544

RESULT 4
AAB69060 standard; Protein; 547 AA.
ID AAB69060
XX
AC AAB69060;
XX
DT 18-APR-2001 (first entry)
XX
DE Pseudomonas aeruginosa groEL protein sequence.
XX
XX Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
KW diagnosis; detection; infection; immune response.
XX
OS Pseudomonas aeruginosa.
XX
XX WO200102577-A1.
XX
XX 11-JAN-2001.
XX
XX 03-JUL-2000; 2000WO-GB02554.
XX
XX 01-JUL-1999; 99GB-0015419.
XX
XX (PROV-) PROVALIS UK LTD.
XX
XX Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PUH, Wilkinson MC;
XX WPI; 2001-080988/09.
DB N-PSDB; AAF32452.
XX
XX Antigenic Pseudomonas aeruginosa proteins, useful in the detection

PT and/or diagnosis of P. aeruginosa infections and for producing vaccines
 PT against P. aeruginosa -
 XX Claim 3; Fig 25; 129pp; English.

XX The present invention describes antigenic Pseudomonas aeruginosa
 CC proteins (P1). The P. aeruginosa proteins have antibacterial activity
 CC and can be used in vaccines and as antagonists. The proteins or their
 CC fragments, or antibodies are useful in the detection and/or diagnosis
 CC of P. aeruginosa. They are also useful for producing a vaccine and
 CC inducing an immune response against P. aeruginosa infection. An agent
 CC capable of antagonising, inhibiting or otherwise interfering with the
 CC function or expression of P1 are useful in the manufacture of a
 CC medicament for the treatment or prophylaxis of P. aeruginosa infections.
 CC The present sequence represents a specifically claimed P. aeruginosa
 CC GroEL protein sequence from the present invention.

XX Sequence 547 AA;

Query Match 67.4%; Score 1825.5; DB 22; Length 547;
 Best Local Similarity 66.8%; Pred. No. 4.3e-112;
 Matches 367; Conservative 75; Mismatches 102; Indels 5; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGPKGRNVVIEKSGSPVITKDGVSVAKEI 60
 DB 1 MAAKEVKFGDSARKMLVGVNVLADAVKATLGPGRNVVLDKSGFAPTITKDGVSVAKEI 60
 QY 61 ELEDKFNMGAMQVKEVAPKTSIDTADGCTTTATVLAQAIYREGVKLVAAGRNPMATKRG 120
 DB 61 ELKDFENMGAMQVKEVAPKTSIDTADGCTTTATVLAQAIYREGVKLVAAGRNPMATKRG 120
 QY 121 DRVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEEA 180
 DB 121 DRVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEEA 180
 QY 181 KGLTTLDDVVEGMKFDGRLSPFYVTNPCKMVCEDNPNVILCNKKITSMKMLPILQEV 240
 DB 181 SGLNELSVVEGMKFDGRLSPFYVTNPCKMVCEDNPNVILCNKKITSMKMLPILQEV 240
 QY 241 AKVNRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEA 300
 DB 241 AKAGRELLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEA 300
 QY 301 IPEDRGIKLENVSLSGTAKRVVVDKENTTIIVDGAGKSEDIKARVKQIRAQIETSSDY 360
 DB 301 ISEVGLLEGATLEHLGNKRVVINKENTTIIDGAGVQADIEARVLQIRKQIETTSY 360
 QY 361 DREKLERLAKLVGGVAVIHVGAATETEMKEKRDVEDALNATRAAEEGIVPGGTAFV 420
 DB 361 DREKLERLAKLVGGVAVIHVGAATEVEKKEKARVEDALHATRAAEEGIVPGGTAFV 420
 QY 421 RSIVKLVDDIKPADDDELALGNIIRSLRPLQIAANAGYEGSVIVVEKREPKDGFNFA 480
 DB 421 RALQAEIKLVGGVAVIHVGAATEVEKKEKARVEDALHATRAAEEGIVPGGTAFV 480
 QY 481 ASGEYEDLLKAGVIDPKVTRIALQNAASVASLTLTTECAIAEKPKKMDMPGG--GM 538
 DB 481 ATGVYDMLIENGILDPKVTSLQAAASIGGLMITTEAMVAIVE---DKPANGMPDM 537
 QY 539 GGMGGMDGM 547
 DB 538 GGMGGMGGM 546

RESULT 5
 AAR67381
 XX AAR67381 standard; Protein; 547 AA.
 AC AAR67381;
 XX
 DT 25-MAR-2003 (updated)
 DT 22-JUN-1995 (first entry)
 XX

DE L. pneumophila HtpB gene product.
 XX
 KW Urease; immunogen; vaccine; diagnostic; heat shock protein; HSP;
 KW GroEL-like protein; Helicobacter felis.
 XX
 OS Legionella pneumophila.
 XX
 PN WO9426901-A1.
 XX
 PD 24-NOV-1994.
 XX
 XX 19-MAY-1994; 94WO-EP01625.
 XX
 PR 19-MAY-1993; 93EP-0401309.
 PR 19-NOV-1993; 93WO-EP03259.
 XX
 XX (INRM) INST NAT SANTE & RECH MEDICALE.
 PA (INSP) INST PASTEUR.
 XX
 XX Ferrero R, Labigne A, Suerbaum S, Thiberge J;
 XX WPI; 1995-006797/01.
 DR
 XX DNA from Helicobacter pylori and Helicobacter felis - used to
 PT develop prods. for detection, treatment and prevention of
 PT Helicobacter infection
 XX
 PS Disclosure; Fig. 7A(i-vii); 169pp; English.
 XX
 CC The sequence of the Helicobacter pylori heat shock protein A
 CC (given in AAR67374) was compared to that of other GroEL-like
 CC proteins from Legionella pneumophila (AAR67381), Escherichia coli
 CC (AAR67382), Chlamydia psittaci (AAR67383), Mycobacterium leprae
 CC (AAR67384) and human mitochondrial protein P1 (AAR67385), and regions
 CC of homology were identified.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 547 AA;

Query Match 67.2%; Score 1820.5; DB 16; Length 547;
 Best Local Similarity 66.6%; Pred. No. 9.1e-112;
 Matches 363; Conservative 82; Mismatches 97; Indels 3; Gaps 2;

QY 3 SKEILFDKAREKLSRGVDKLANAVKVTGPKGRNVVIEKSGSPVITKDGVSVAKEIEL 62
 DB 2 AKELRFGDARLOMLAGVNALADAVVTGPRGRNVVLEKSYGAPVTVKDGVSAKEIEF 61
 QY 63 EDKFNMGAMQVKEVAPKTSIDTADGCTTTATVLAQAIYREGVKLVAAGRNPMATKRGIDK 122
 DB 62 EHRFNMGMAMQVKEVAPKTSIDTADGCTTTATVLAQAIYREGVKLVAAGRNPMATKRGIDK 121
 QY 123 AVVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEAKG 182
 DB 122 AVLAVTKELQAMSKPKCKSKAIAQVGTISANSDEAIGAIIEAMEKVGEGVITVEDGNG 181
 QY 183 LETTLDVVEGMKFDGRLSPFYVTNPCKMVCEDNPNVILCNKKITSMKMLPILQEVAK 242
 DB 182 LENELVVEGMKFDGRLSPFYVTNPCKMVCEDNPNVILCNKKITSMKMLPILQEVAK 241
 QY 243 VNRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEAIF 302
 DB 242 SGRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEAIF 301
 QY 303 ERGKLENVSLSGTAKRVVVDKENTTIIVDGAGKSEDIKARVKQIRAQIETSSDYDR 362
 DB 302 EEIGKSLGATLEDLGSARKIIVTDE--TTIIDEGKATIEINARIAQIRAOQMEETTSYDR 360
 QY 363 EKLOERLAKLVGGVAVIHVGAATEVEKKEKRDVEDALNATRAAEEGIVPGGTAFVRS 422
 DB 361 EKLOERLAKLVGGVAVIHVGAATEVEKKEKRDVEDALHATRAAEEGIVAGGVALLIRA 420
 QY 423 IKVLDIDKPADDELALGNIIRSLRPLQIAANAGYEGSVIVVEKREPKDGFNFAAS 482

Db 421 QKALDSLKDDQDQNGINILRRATESPMRQIVTNAGYEASVNVNVAEHKDNKYNFNAAT 480
 QY 483 GEYEDLIKAGVIDPKKVTTRIALQNAASVASLLLTTECAIAEKPEPKDMPFGGGMGMG 542
 Db 481 GEYGDWEMGIIDPTKTRMALQNAASVASLMLTTECMVADL--PKKEGVGAGDMGMGMG 538
 QY 543 GMDGM 547
 Db 539 GMGMG 543

RESULT 6
 AAB69061
 ID AAB69061 standard; Protein; 574 AA.
 XX
 AC AAB69061;
 XX
 DT 18-APR-2001 (first entry)
 XX
 DE pTrcHisB expression vector groEL fusion protein sequence.
 XX
 KW Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
 KW diagnosis; detection; infection; immune response.
 XX
 OS Pseudomonas aeruginosa.
 OS Synthetic.
 XX
 PN W0200102577-A1.
 XX
 PD 11-JAN-2001.
 XX
 PF 03-JUL-2000; 2000WO-GB02554.
 XX
 PR 01-JUL-1999; 99GB-0015419.
 XX
 PA (PROV-) PROVALIS UK LTD.
 XX
 PI Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PWH, Wilkinson MC;
 XX
 DR WPI; 2001-080988/09.
 DR N-PSDB; AAF32453.
 XX

Antigenic Pseudomonas aeruginosa proteins, useful in the detection and/or diagnosis of P. aeruginosa infections and for producing vaccines against P. aeruginosa -
 XX
 PS Claim 3; Fig 26; 129pp; English.
 XX

The present invention describes antigenic Pseudomonas aeruginosa proteins (Pl). The P. aeruginosa proteins have antibacterial activity and can be used in vaccines and as antagonists. The proteins or their fragments, or antibodies are useful in the detection and/or diagnosis of P. aeruginosa. They are also useful for producing a vaccine and inducing an immune response against P. aeruginosa infection. An agent capable of antagonising, inhibiting or otherwise interfering with the function or expression of Pl are useful in the manufacture of a medicament for the treatment or prophylaxis of P. aeruginosa infections. The present sequence represents a specifically claimed pTrcHisB expression vector groEL fusion protein sequence from the present invention.
 XX
 SQ Sequence 574 AA;

Query Match 67.2%; Score 1819; DB 22; Length 574;
 Best Local Similarity 66.7%; Pred. No. 1.2e-111;
 Matches 365; Conservative 75; Mismatches 103; Indels 4; Gaps 1;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGPKGRNVIEKSPGVITKDGYSVAKEI 60
 Db 31 MAKEVFGDSARKMLVGNVLADAVKATGPKGRNVLDKSPGAPITTKDGYSVAKEI 90
 QY 61 ELEDKFNMGAMQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAIKRGI 120

Db 91 ELKDKFENMGAQLVKDVASKANDAAAGDGTATVLAQAIYNEGLKAVAAGMNPMDLKRGI 150
 QY 121 DKAVAVTKELSDITKPTRDOKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGGVITVBEA 180
 Db 151 DKATVAIVAAIKELAKPCADTKAIAQVGTISANDESIGQIIAEAMERKVGEGVITVBEG 210
 QY 181 KGLETTLDVWEGMKFDRGYLSPYFVTNPEKVVCELDNPIILCNEKKITSMKMDLPIILQV 240
 Db 211 SGLNELSVVEGMQFDRGYLSPYFNKPDPTMAAELDSPLLLLVDDKKISNIREMLPVLBAV 270
 QY 241 AKVNRPLIIIAEDVEGEALATLVVNLKRGALQVAVKAPGFGERRKAMLEDIAIITGGEA 300
 Db 271 AKAGREPLIIIAEDVEGEALATLVVNMREGIVKVAAVKAPGFGDRKAMLQDIAIITGGTV 330
 QY 301 IFEDRGIKLENVSLSGTAKRVVIDKENTTIVGAGKSEDIKARVKQIRAQIETSSDY 360
 Db 331 ISEEVGLSLEGATLEHLGNKRVVINKENTTIIDGAGVQADIEARVLQIRKQIETTSY 390
 QY 361 DREKLOERLAKLVGVAVIHVGAATETEMKEDKDRVEDALNATRAAAVEGIVPGGGTAFV 420
 Db 391 DREKLOERLAKLVGVAVIKVGAATEVEMKEDKARVEDALHATRAAAVEGVVPGGGVALV 450
 QY 421 RSIKVLDDIKPADDDLAGLMIIRRSLEEPURQIAANAGYEGSIVVEKVRPDKDGFEN 480
 Db 451 RALQAIIEGLKGDNEEQNVGIALRRAVESPLRQIVANAGDEPSVVVDKVGKSGNYGNA 510
 QY 481 ASGEYEDLIKAGVIDPKKVTTRIALQNAASVASLLLTTECAIAEKPEPKDMPFGGGMG 540
 Db 511 ATGYGDMIEGILDPAKVTTRSAIQAASISGLMITTEAMVAEIVEDKPA---GMPD 566
 QY 541 MGGMDGM 547
 Db 567 MGGMGGM 573

RESULT 7
 AAY23915
 ID AAY23915 standard; Protein; 545 AA.
 XX
 AC AAY23915;
 XX
 DT 22-SEP-1999 (first entry)
 XX
 DE Amino acid sequence of a heat shock protein.
 XX
 KW Heat shock protein; Hsp; immune response; immunological carrier;
 KW cancer control; tumour; sarcoma; cancer; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 EN W09935270-A1.
 XX
 PD 15-JUL-1999.
 XX
 PF 29-DEC-1998; 98WO-CA01203.
 XX
 PR 31-DEC-1997; 97US-0001737.
 XX
 PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 XX
 PI Mizzen L, Wisniewski J;
 XX
 DR WPI; 1999-430397/36.
 XX

New nucleic acid encoding heat shock protein-60 from Streptococcus, useful in vaccines, as carriers for other immunogens, as anticancer agents and for diagnosis
 XX
 PS Disclosure; Fig 10A-E; 176pp; English.
 XX

AAY23905-30 represent heat shock proteins (Hsps). The specification describes Streptococcal Hsps, designated Hsp60. These proteins, their fragments, variants and fusion proteins, are used to elicit or enhance

CC an immune response against Streptococcus, and to elicit a similar
 CC response to a target antigen fused to the protein. Unlike other
 CC immunological carriers, Hsp60 proteins are not immunosuppressive so
 CC provide an increased response to any conjugated or fused antigen. Also,
 CC where used for cancer control, they lack the side effects associated
 CC with endotoxins. They can also be used to detect specific antibodies
 CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
 CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
 CC polynucleotide is used for recombinant production of the protein, as
 CC a source of primers and probes for detecting streptococci in standard
 CC hybridization/amplification assays, and therapeutically in gene
 CC therapy vectors.

XX
 SQ Sequence 545 AA;

Query Match 67.0%; Score 1814.5; DB 20; Length 545;
 Best Local Similarity 66.7%; Pred. No. 2.2e-111;
 Matches 363; Conservative 77; Mismatches 103; Indels 1; Gaps 1;

QY 1 MASKEILFDAKAREKLSRGVDKLANAVKVTGLGKGRNVVIEKSGSPVITKDGVSVAKEI 60
 DB 1 MAAKVQFGNEVRQKRWNGVNIILANAVRTLGPGRNVVDRAFGGPHITKDGVTVAKEI 60
 QY 61 ELEDKPFENMGAVQWKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAGRNPMIAKRG 120
 DB 61 ELKDKPFENMGAVQWKEVASKTNDVAGDGTITATVLAQSIIVAECKVYTAGMNPDTLKRGI 120
 QY 121 DRKAVAVTKELSDITKPTDRQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGGVITVEEA 180
 DB 121 DRKAVAVTKELSDITKPTDRQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGGVITVEEDG 180
 QY 181 KGLTETLDVVEGKMPDRGYLSPYFTNPEKMWCELDNPKYLLCNEKKITSMKMLPILQEV 240
 DB 181 KLENELDVVEGKMPDRGYLSPYFTNPEKMWCELDNPKYLLCNEKKITSMKMLPILQEV 240
 QY 241 AKVNRPLLIITAEVDEGEALATLVNKLGRALQVAVKAPGFGERRKAMLEIDIALTGEEA 300
 DB 241 AKASRPLLIITAEVDEGEALATLVNKLGRALQVAVKAPGFGERRKAMLEIDIALTGTTV 300
 QY 301 IFEDRGKILNLSLGTAKRVVDKENTTIVDGAGKSEDIKARVKQIPRAQIETESSDY 360
 DB 301 ISEVGLSLKATLDLQAKRIEIKENTTIIDFGDAAQIEARVAEIRQIETATSDY 360
 QY 361 DRKLOERKLVGVAVTHVGAATETEMKEKKORVEDALNATRAAVEGIVPGGTFAPV 420
 DB 361 DRKLOERKLVGVAVTHVGAATETEMKEKKORVEDALNATRAAVEGIVPGGTFAPV 420
 QY 421 RSIKVLDDIKPADDDLAGLNIIRRSLEELPLQIAANAGYEGSIVVEKVRPKDGFQFNA 480
 DB 421 RARAALNLTGNDQDAGVQIVLRAVESPLRQIVANAGGSPSVVNVKLEGGNYGNA 480
 QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEPKDMPFGGMMG 540
 DB 481 GSGEYEDMIEMGVLDPAKVTRSALQHAASLAGLMLTDCMIAETPEEKPAFDMGGMG 539
 QY 541 MGCM 544
 DB 540 MGCM 543

RESULT 8

ID AAY75747
 ID AAY75747 standard; Protein; 544 AA.

XX AAY75747;

XX AC
 XX AC
 DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 982 protein sequence SEQ ID NO:2966.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

XX Neisseria meningitidis.
 OS WO9957280-A2.
 XX 11-NOV-1999.
 XX 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 DR N-PSDB; AA254509.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 2; Page 1388; 1453pp; English.
 XX AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA254594
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254594 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX SQ Sequence 544 AA;

Query Match 66.9%; Score 1812.5; DB 21; Length 544;
 Best Local Similarity 66.5%; Pred. No. 3e-111;
 Matches 362; Conservative 78; Mismatches 103; Indels 1; Gaps 1;

QY 1 MASKEILFDAKAREKLSRGVDKLANAVKVTGLGKGRNVVIEKSGSPVITKDGVSVAKEI 60
 DB 1 MAAKVQFGNEVRQKRWNGVNIILANAVRTLGPGRNVVDRAFGGPHITKDGVTVAKEI 60
 QY 61 ELEDKPFENMGAVQWKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAGRNPMIAKRG 120
 DB 61 ELKDKPFENMGAVQWKEVASKTNDVAGDGTITATVLAQSIIVAECKVYTAGMNPDTLKRGI 120
 QY 121 DRKAVAVTKELSDITKPTDRQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGGVITVEEA 180
 DB 121 DRKAVAVTKELSDITKPTDRQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGGVITVEEDG 180
 QY 181 KGLTETLDVVEGKMPDRGYLSPYFTNPEKMWCELDNPKYLLCNEKKITSMKMLPILQEV 240
 DB 181 KLENELDVVEGKMPDRGYLSPYFTNPEKMWCELDNPKYLLCNEKKITSMKMLPILQEV 240
 QY 241 AKVNRPLLIITAEVDEGEALATLVNKLGRALQVAVKAPGFGERRKAMLEIDIALTGEEA 300
 DB 241 AKASRPLLIITAEVDEGEALATLVNKLGRALQVAVKAPGFGERRKAMLEIDIALTGTTV 300

QY 301 IFEDRGIKLENVSLSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
 Db ISEEVGLSIEKATLDDLGOAKRIEIGKENTTIIIDFGDAAQIEARVAEIROIETATSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGAATEMKEKKORVEDALNATRAAAVEEGIVPGGGTAFV 420
 Db DKEKLOERLAKLVGGVAVIKVGAATEVEMKEKKORVEDALHATRAAAVEEGVAGGVALL 420
 QY 421 RSIKVLDIDKPADDDELAGLNIIIRSLLEPLRQIAANAGYEGSVVEKVPKDGFGENA 480
 Db RARAENLHTGNADQDAGVOIVLRAVESPLRQIVANAGGSPSVVNVKLGKNGYGYNA 480
 QY 481 ASGEYEDLIKAGVIDPKKVTTRIALQNAASVASLTLTTECAIAEKPEPKOMPMPGGG 540
 Db GSGEYGDWIEGVLDPAKVTRSAHQHAASITAGMLTTDCMTAEIPEKPAFV 539
 QY 541 MGGM 544
 Db 540 MGGM 543

RESULT 9
 ABP77245
 ID ABP77245 standard; Protein; 544 AA.
 AC ABP77245;
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae amino acid sequence SEQ ID 1020.
 KW Antibacterial; infection; vaccine; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 FN WO200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-IB02069.
 XX
 PR 12-FEB-2001; 2001GB-0003424.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizza M, Massignani V, Monaci E;
 XX
 DR WPI; 2003-058415/05.
 DR N-PSDB; ABZ38215.
 XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 XX
 PS Disclosure; Page 262; 815pp; English.
 XX
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX
 SQ Sequence 544 AA;

Query Match 66.7%; Score 1805.5; DB 24; Length 544;
 Best Local Similarity 66.4%; Pred. No. 8.8e-111;
 Matches 361; Conservative 78; Mismatches 104; Indels 1; Gaps 1;

QY 1 MASKEIIFDAKAREKLSRGVDKLANAVKVTGLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
 Db 1 MAADKDVQFGNEVRQKMWNGVNLANAVRVTGLGPKGRNVVVDRAFGGPHITKDGVTVAKEI 60

QY 61 ELEDKFNMGQAMQVKEVAPKTSIDTAGDGTATTATVLAQAIYEGVKLYAAGRNPAIKRGI 120
 Db ELKDKFNMGQAMQVKEVASKTNDVAGDGTATTATVLAQSIYAEGRMKYVTAGNPTDLKRGI 120
 QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAABAMAKVGGVITVBEA 180
 Db DKAAALVEELKNIAKPCDTSKEIAQVGSISANSDEQVGAIIAEMKVKEGVITVBDG 180
 QY 181 KGLTTLDVVEGMPKDRGYLSPYFVTNPKVWCELDNPIILCNKKTITSMKMDLPILEQV 240
 Db KSLNELDVVEGMPKDRGYLSPYFINDAEKQIAGLDNPFVLLFDKKTISNIRDLPLVLEQV 240
 QY 241 AKVNRPLLIITAEVDVEGALATLVNKLKRGALOVVAVKAPGGERKAKMLEDIALLTGSEA 300
 Db AKASRPLLIITAEVDVEGALATLVNKRIRGLIKTVAVKAPGGERKAKMLEDIALLTGSEV 300
 QY 301 IFEDRGIKLENVSLSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
 Db ISEEVGLSLEKATLDDLGOAKRIEIGKENTTIVDGFGDAAQIEARVAEIROIETATSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGAATEMKEKKORVEDALNATRAAAVEEGIVPGGGTAFV 420
 Db DKEKLOERLAKLVGGVAVIKVGAATEVEMKEKKORVEDALHATRAAAVEEGVAGGVALL 420
 QY 421 RSIKVLDIDKPADDDELAGLNIIIRSLLEPLRQIAANAGYEGSVVEKVPKDGFGENA 480
 Db RARAENLHTGNADQDAGVOIVLRAVESPLRQIVANAGGSPSVVNVKLGKNGYGYNA 480
 QY 481 ASGEYEDLIKAGVIDPKKVTTRIALQNAASVASLTLTTECAIAEKPEPKOMPMPGGG 540
 Db GSGEYGDWIEGVLDPAKVTRSAHQHAASITAGMLTTDCMTAEIPEKPAFV 539
 QY 541 MGGM 544
 Db 540 MGGM 543

RESULT 10
 AAE22070
 ID AAE22070 standard; Protein; 544 AA.
 AC AAE22070;
 XX
 DT 25-JUL-2002 (first entry)
 XX
 DE Francisella tularensis GroEL protein.
 KW Francisella tularensis
 XX
 OS Heat shock protein; hsp60; GroEL protein; therapy; infection; antibiotic.
 XX
 FN WO200218600-A1.
 XX
 PD 07-MAR-2002.
 XX
 PF 24-AUG-2001; 2001WO-GB03835.
 XX
 PR 26-AUG-2000; 2000GB-0021008.
 XX
 PA (MINA) UK SEC FOR DEFENCE.
 XX
 PI Titball RW, Green M, Hartley MG, Mack KA, Sjoestedt A;
 XX
 DR WPI; 2002-362179/39.
 XX
 PT Novel polypeptide comprising a component of heat shock protein, Hsp60
 PT obtainable from Francisella tularensis, useful for preventing or
 PT treating infection caused by Francisella tularensis in an animal -
 XX
 PS Disclosure; Fig 1; 26pp; English.
 XX
 CC The invention relates to a polypeptide comprising at least one

CC component of an heat shock protein 60 (hsp60) obtainable from
 CC Francisella tularensis, and in particular the GroEL protein or its
 CC fragment, and a nucleic acid sequence encoding the polypeptide. The
 CC polypeptide and polynucleotides of the invention are useful for
 CC preparing a medicament and for preventing or treating infection
 CC caused by F. tularensis in an animal. The polynucleotide of the
 CC invention is administered to an animal in a form in which it is
 CC expressed in situ. The present sequence is Francisella tularensis
 CC GroEL protein.

XX SQ Sequence 544 AA;

Query Match 65.3%; Score 1769.5; DB 23; Length 544;

Best Local Similarity 64.9%; Pred. No. 21e-108;

Matches 353; Conservative 77; Mismatches 113; Indels 1; Gaps 1;

QY 1 MASKEILFPAKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSGSPVITKDGVSVAKEI 60
 DB 1 MAAKQVLFSDARAKMLDGVNTLANAVKVTLPKGRNVVLDKSGFGAPTITKDGVSVAKEI 60
 QY 61 ELEDKEFNMGCAQWKEVAPKTSIADGGTTATVLAQAIYREGVKLVAAGRNPMATKRG 120
 DB 61 ELEDKEFNMGCAQWKEVAPKTSIADGGTTATVLAQALTEGLKAVTAGMNPMDLKRG 120
 QY 121 DRVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAAMAKVKGKGVITVEEA 180
 DB 121 DRATRLVBEELKALSPKSPDPSIEQVGTISANSDDTTIGNIIAAMAKVKGKGVITVEEG 180
 QY 181 KGLTTLDDVVEGKMFDRGYLSPYFTNPERKWCENLNPYILCNEKITSKMDLPVLEQV 240
 DB 181 KGFEDLDDVVEGKMFDRGYLSPYFTNPERKWCENLNPYILCNEKITSKMDLPVLEGV 240
 QY 241 AKVNRPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEDAIITGGEA 300
 DB 241 SKSGRALLIIAEDVEGEALATLVNKNMRGVVCTVKAPGFGERRKAMLEDAIITGATF 300
 QY 301 IFEDRIGIKLENVSLSIGTAKRVVVDKENTITVDGAGKSDIKARVKQIRAEETSSDY 360
 DB 301 VSEDLMSKLEETNWEHLGTASRVQVTKDNTIIDGAGEKAIKRNIVIKANAEANSY 360
 QY 361 DREKLERIAKLVGGVAVIHVGATETEMKEDKVEDALNATRAAVEGIVPGGTAFV 420
 DB 361 DREKLERIAKLVGGVAVIKVGAETAEKEDKVEDALNATRAAVEGIVAGGGVALI 420
 QY 421 RSLKVLDDIKPADDELALGNIIRSLLEPLROIAANAGYEGSVVVEKVPKDGCFNA 480
 DB 421 RAQKALDGLTGENDQNHGIALRLKAIEAPLRQIVSNAGGESSVNVQKANOQNGYNA 480
 QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPPKMDPMPGGMGG 540
 DB 481 ANDTYGDMVEMGILDPKTVTRSLQHAASLAGLMTTEAMIGELKEAAPMPM-GGGMGG 539
 QY 541 MGQM 544
 DB 540 MPQM 543

RESULT 11

AA75745

ID AA75745 standard; Protein; 544 AA.

XX AAY75745;

AC AAY75745;

XX 21-MAR-2000 (first entry)

XX Neisseria gonorrhoeae ORF 982 protein sequence SEQ ID NO:2962.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.

XX Neisseria gonorrhoeae.

OS

XX

PN MO957280-AA2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

XX Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AA254507.

XX Novel Neisserial polypeptides predicted to be useful antigens for

XX PT vaccines and diagnostics -

XX Claim 2; Page 1385; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941

XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent

XX PCR primers used in the exemplification of the present invention. The

XX polypeptides, the polynucleotides, antibodies and compositions of

XX the invention can be used as vaccines, as diagnostic reagents, and as

XX immunogenic compositions. The polypeptides can be used in the

XX manufacture of medicaments for treating or preventing infection due to

XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the

XX presence of Neisseria bacteria, or to raise antibodies. They may also

XX be used to screen for agonists or antagonists, which may themselves

XX have use as antibacterial agents. The polynucleotides of the invention

XX may also be used in gene therapy protocols.

XX SQ Sequence 544 AA;

Query Match 64.8%; Score 1750.5; DB 21; Length 544;

Best Local Similarity 64.9%; Pred. No. 3.7e-107;

Matches 353; Conservative 78; Mismatches 112; Indels 1; Gaps 1;

QY 1 MASKEILFPAKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSGSPVITKDGVSVAKEI 60

DB 1 IASQNRFRNFQKRVNGVNIIPADWALGAKGRNVVVDRAFGGPHITKDGVSVAKEI 60

QY 61 ELEDKEFNMGCAQWKEVAPKTSIADGGTTATVLAQAIYREGVKLVAAGRNPMATKRG 120

DB 61 ELKDKFENMGCAQWKEVAPKTSIADGGTTATVLAQSIVAEGMKVYTAGMNPDLKRG 120

QY 121 DRVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAAMAKVKGKGVITVEEA 180

DB 121 DRVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAAMAKVKGKGVITVEEG 180

QY 181 KGLTTLDDVVEGKMFDRGYLSPYFTNPERKWCENLNPYILCNEKITSKMDLPVLEQV 240

DB 181 KSLNEDLVVEGKMFDRGYLSPYFTNPERKWCENLNPYILCNEKITSKMDLPVLEQV 240

QY 241 AKVNRPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEDAIITGGEA 300

DB 241 AKASRPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEDAIITGGEA 300

QY 301 IFEDRIGIKLENVSLSIGTAKRVVVDKENTITVDGAGKSDIKARVKQIRAEETSSDY 360

DB 301 ISEVGLSLKATLDDLGQTKRIEIGENTTVIDGFGDAAQIEARVAEIRQIETATSDY 360

RESULT 14

RESULT 15

AAW95001
ID AAW95001 standard; Protein; 548 AA.

AC AAW95001;

DT 12-MAY-1999 (first entry)

DE WO9902989 Seq ID 10.

KW Chaperone; molecular chaperone; groEL gene.

OS *Escherichia coli*.

PN WO9902989-A1.

21--JAN-1999.

AA
PF 10-JUL-1998: 98WO-GB02042.

PR 28-AUG-1997; 97GB-0018261.

FR 10-JUL-1997; 9/GB-0014582.
XX

FA (MEDI-) MEDICAL RES COUNCIL.
XX

PI Chateallier J, Fersht A;
XX

DR WPI; 1999-121113/10.
XX

PT In vivo assay for the detection of chaperone fragment activity -
PT useful for, e.g. complementing a molecular chaperone defect in vivo
PT

PS Disclosure; Page 52-55; 64pp; English.

The invention provides a method f providing chaperone activity in vivo, by administering to a cell a fragment of a molecular chaperone which has in vivo activity. Also provided is a method of determining whether a fragment of a molecular chaperone is active in vivo by: (i) providing a cell with a deficient molecular chaperone activity; (ii) administering the fragment to the cell; and (iii) determining whether the fragment complements the deficient endogenous molecular chaperone. The method can determine which fragments of the molecular chaperone can be used to complement a mutant or deficient molecular chaperone in vivo. The method allows an in vivo assessment of chaperone activity as compared to previous tests which only measured activity in vitro.

Sequence 548 AA;

Query Match	64.5%;	Score 1746;	DB 20;	Length 548;
Best Local Similarity	64.5%;	Pred. No. 7.3e-107;		
Matches 354;	Conservative	74;	Mismatches 117;	Indels 4;
Gaps 2;				

Qy 1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFSGSPVITKDGVSVAKEI 60

Db

1 MAAKDVKEGNDARVKMLRGVNVVLADAVKVTLGPKGRNVVLDKSFAPTITKDGVSVAREI 60

QY 61 ELEDKFENMGAQMVKVAPKTSIDAGDGTATTATVLAQAIYREGVKLVAAGRNPMATKRG I 120

D**b** 61 ELEDKFENMGAQMVKVASKANDAAGDGTATTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120

QY 121 DKAVAVTKELSDITKPTRDQKEIAQVGTISANSDDTIGNIIAEAMAKVGKGGVITVEEA 180

D**b** 121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLI AEAMDKVKGEGVITVEDG 180

QY 181 KGLETTLDVVEGMKFDGRGYSPLYFTNPEKMVCELDNPYILCNEKKITSMKMDMLPILEQV 240

Db 181 TGLQDEL DVVEGMQFDRGYLSPYFINKPETGAVELESFFILLADKKISNIREMLPYLEAV 240

QY 241 AKVNRPLIIAEDVEGEALATLVNKLRGALOVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db
241 AKAGKPLIIIAEDVEGEALATLVNTMRGIVKVAAVKAPGEGDPRKAMLODIATLTGGTV 300

QY 301 IFEDRGIKLENVSLSSLTAKRVVIDKENTTIVDGAGKSEDIKARVKOIRAOIETSSDY 360

	: {:: : : : : : : : :	
Db	301 I SEIGMELEKATLEDLGQAKRVVINKDTTIIIDVGEEAIIQGVAIQROIQIEATS DY	360
Qy	361 DREKLQRLAKLVGGVAIVHVGAAATEMKEKKORVEDALNATRAAVEEIVPGGTA FV	420
Db	361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGV ALI	420
Qy	421 RSIKVLDDIPADDELGLAGNIIRSELEEPLRQIAANAGYEGSI VVEKVREPKDGFNA	480
Db	421 RVASKLADLRGQEDNQNVGIKVALURAMEAPUQRIVLNGCEPESVVANTVKGDGN YGNA	480
Qy	481 ASGEYEDLIKAGVIDPKKVTRIALQNASVASLLLTTECAJAEKPEPKDMMPGC -- GGM	538
Db	481 ATBEYGNMIDMGILDPDKVTRSALQYASVASGLMITTECMVTDL--PRNDAADLGAAG M	538
Qy	539 GGMGMDGM 547 	
Db	539 GGMGMDGM 547 	

Search completed: January 28, 2004, 13:15:54
Job time : 49 secs

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:15:00 ; Search time 22 Seconds
(without alignments)
1053.925 Million cell updates/sec

Title: US-09-077-574A-2

Perfect score: 2708

Sequence: 1 MASKEILFDKAREKLSRGV.....KDMPPGGGGMGGMGMDGMY 548

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1844	68.1	548	2	US-08-467-822-31
2	1844	68.1	548	3	US-08-432-697-31
3	1844	68.1	548	3	US-08-466-248-31
4	1820.5	67.2	551	4	US-09-252-991A-24209
5	1768.5	65.3	550	4	US-09-328-352-5508
6	1741	64.3	548	2	US-08-467-822-32
7	1741	64.3	548	3	US-09-472-971-3
8	1741	64.3	548	3	US-08-432-697-32
9	1741	64.3	548	3	US-08-466-248-32
10	1723.5	63.6	546	3	US-08-470-260-6
11	1723.5	63.6	546	3	US-08-471-491-6
12	1723.5	63.6	546	3	US-08-466-662-6
13	1723.5	63.6	546	3	US-08-256-847C-1
14	1723.5	63.6	546	4	US-08-256-847C-7
15	1721	63.6	544	2	US-08-467-822-33
16	1721	63.6	544	3	US-08-432-697-33
17	1721	63.6	544	3	US-08-466-248-33
18	1716.5	63.4	547	4	US-08-461-722-2
19	1716.5	63.4	547	4	US-08-336-251-2
20	1716.5	63.4	547	4	US-09-468-041-2
21	1716.5	63.4	547	5	PCT-US94-06362-2
22	1708	62.9	545	4	US-09-198-452A-153
23	1702	62.9	545	2	US-08-467-822-30
24	1702	62.9	545	3	US-08-432-697-30
25	1702	62.9	545	3	US-08-466-248-30
26	1683.5	62.2	641	4	US-09-613-303-51
27	1597	59.0	541	2	US-08-997-080-160

Sequence 160, Appl
Sequence 160, Appl
Sequence 160, Appl
Sequence 160, Appl
Sequence 4, Appl
Sequence 17, Appl
Sequence 29, Appl
Sequence 21, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4705, Ap
Sequence 34, Appl
Sequence 19, Appl
Sequence 34, Appl
Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-08-467-822-31

; Sequence 31, Application US/08467822

; Patent No. 5843460

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; FILING DATE: 19-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0137-02000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 31:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 548 amino acids

; TYPE: amino acid

US-08-432-697-32

Query Match 64.3%; Score 1741; DB 3; Length 548;
Best Local Similarity 64.3%; Pred. No. 1e-143;
Matches 353; Conservative 74; Mismatches 118; Indels 4; Gaps 2;

QY 1 MASKELFPAKAREKLSRGVDKLANAVKVTGLGPKGNVIEKSGSPVITKDGVSVAKEI 60
DB 1 MAADKYKFGNDARVKMLRGVNLADAVKVTGLGPKGNVLDKSGFGAPTITKDGVSVAKEI 60

QY 61 ELEDKFNMGAWQVKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAAGRNPMIAIKRGI 120
DB 61 ELEDKFNMGAWQVKEVASKANDADGTTTATVLAQAIITGLKVAAGMNMPLDKRGI 120

QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANDTTIGNIIAEMAKVKGGVITVEEA 180
DB 121 DRAVTAABEELKALSVPKSDSKAIAQVGTISANDTTIGNIIAEMAKVKGGVITVEEDG 180

QY 181 KGLTTLDVVEGKMFEDRGYLSPVFTNPEKMWCELDNPVILCNEKTTSMKMDLPVLEOV 240
DB 181 TGLQDELVDVVEGQFDRGYLSPVFTNPEKMWCELDNPVILCNEKTTSMKMDLPVLEAV 240

QY 241 AKVNRPLLIITAEVDEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDIALLTGGEA 300
DB 241 AKAGKPLLIITAEVDEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDIALLTGGEA 300

QY 301 IFEDRGKLENVSLSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
DB 301 ISEIEMELEKATLEDLGQAKRVVINKDITTIIDVGEEAAIQGRVAQIRQIIEEATSDY 360

QY 361 DREKLOERLAKVGVAVIHVGAATETEMKEKDRVEDALNATRAAEEGIVPGGTAFAV 420
DB 361 DREKLOERLAKVGVAVIHVGAATEVEMKEKARVEDALHATRAAEEGIVAGGVALLI 420

QY 421 RSTIKVLDDIKPADDELALAGNIIRRSLEELPQIAANAGYEGSVIVVEKPREKDPGFNA 480
DB 421 RVASKLADLRGQNEQVGLKVALRAMEAPLRQIVLNCGEPSVAVNTVKGDDGNYGNA 480

QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEKMPMPG--GGM 538
DB 481 ATEEYGNMIDMGILDPTKVRSLQYAAVAGLMIITTECMVTDL--PKNDAADLGAAGM 538

QY 539 GGMGMDGM 547
DB 539 GGMGMDGM 547

RESULT 9

US-08-466-248-32
Sequence 32, Application US/08466248
Patent No. 6258359
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,248

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/447,177

FILING DATE: 19-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/432,697

FILING DATE: 02-MAY-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03495.0137-02000

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 548 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-466-248-32

Query Match 64.3%; Score 1741; DB 3; Length 548;

Best Local Similarity 64.3%; Pred. No. 1e-143;

Matches 353; Conservative 74; Mismatches 118; Indels 4; Gaps 2;

QY 1 MASKELFPAKAREKLSRGVDKLANAVKVTGLGPKGNVIEKSGSPVITKDGVSVAKEI 60
DB 1 MAADKYKFGNDARVKMLRGVNLADAVKVTGLGPKGNVLDKSGFGAPTITKDGVSVAKEI 60

QY 61 ELEDKFNMGAWQVKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAAGRNPMIAIKRGI 120
DB 61 ELEDKFNMGAWQVKEVASKANDADGTTTATVLAQAIITGLKVAAGMNMPLDKRGI 120

QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANDTTIGNIIAEMAKVKGGVITVEEA 180
DB 121 DRAVTAABEELKALSVPKSDSKAIAQVGTISANDTTIGNIIAEMAKVKGGVITVEEDG 180

QY 181 KGLTTLDVVEGKMFEDRGYLSPVFTNPEKMWCELDNPVILCNEKTTSMKMDLPVLEOV 240
DB 181 TGLQDELVDVVEGQFDRGYLSPVFTNPEKMWCELDNPVILCNEKTTSMKMDLPVLEAV 240

QY 241 AKVNRPLLIITAEVDEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDIALLTGGEA 300
DB 241 AKAGKPLLIITAEVDEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDIALLTGGEA 300

QY 301 IFEDRGKLENVSLSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
DB 301 ISEIEMELEKATLEDLGQAKRVVINKDITTIIDVGEEAAIQGRVAQIRQIIEEATSDY 360

QY 361 DREKLOERLAKVGVAVIHVGAATETEMKEKDRVEDALNATRAAEEGIVPGGTAFAV 420
DB 361 DREKLOERLAKVGVAVIHVGAATEVEMKEKARVEDALHATRAAEEGIVAGGVALLI 420

QY 421 RSTIKVLDDIKPADDELALAGNIIRRSLEELPQIAANAGYEGSVIVVEKPREKDPGFNA 480
DB 421 RVASKLADLRGQNEQVGLKVALRAMEAPLRQIVLNCGEPSVAVNTVKGDDGNYGNA 480

QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEKMPMPG--GGM 538
DB 481 ATEEYGNMIDMGILDPTKVRSLQYAAVAGLMIITTECMVTDL--PKNDAADLGAAGM 538

QY 539 GGMGMDGM 547
DB 539 GGMGMDGM 547

Qy	3	SKBILFPAKAREKILSRGVDKLANAVKVTLPKGRNVVIEKSFSGSPVITKDGVSVAKEI	62
Db	2	AKBIFKSDSARNLLFEGVRJHDIAVKVTMTGPRGENVLIQKSYGAFSITPKDGVSVAKEI	61
Qy	63	EDRFENGAQWKEVAPKTSIDAGDGTTTATVLAQAATYREGVKLVAAGNPMIAIKRGIDK	122
Db	62	SCPVANNGAOLVKEVASKTADAGDGTTTATVLAYSIFKEGLRMTAGANPIEVKRGMDK	121
Qy	123	AVVAVTKEISDITKPTRDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGGVITVEEAKG	182
Db	122	AABAINELKKAASKYCKGKEITQVATISANSDHNIGKLADAMEKVGKDGVTVEEAKG	181
Qy	183	LETTLDVVEGMKFDGRGYSLPYFVTNPEKMVCELDPYILCNEKKITSMKMDLPILEQVAK	242
Db	182	IEDELVDVEGMQFDRGYSLPYFVTNAEKMTAQDNAVILLTDKKISMMKDILPLEKTMK	241
Qy	243	VNEPLLIIAEDVVEGEALATLVNKLGAOLQVAVKAPFGFERRKAMLEDIAILTGGEAIF	302
Db	242	EGRPLLIIAEDIEGEALTTLVNKLGRGLVIAAVKAPFGFDRRKEMLKDIAILTGGQVIS	301
Qy	303	EDRGIKLENVLSLGTAKRWIDKENTTTVDGAGKSEDIKARVKQIRAQIETEESDDYDR	362
Db	302	EELGLSEUAEVFLGVAGRIVIDKONTTTVDGKGSHDDVKORVAQIKQTIASTISYDK	361
Qy	363	EKLQERLAKLVGVAVITHUGAAATETEMKEKKORVEDALNATRAAEVBEIGVPGGTA	422

Db 362 EKIQLERAKLSGVAVIKVGAASEVEMKEKDRVDALLSATKAAVEEGIVIGGAALIRA 421
QY 423 IKVLDDIKPADDELALAGLNIIRSLLEPLRQIAANAGYEGSIIVKVPKDFGFENAS 482
Db 422 AOKVH--LNLHDEKVGVEIIMRAIKAPLAQIAINAGYDGVVNEVEKEHGFHFNASN 479
QY 483 GEYEDLIKAGVIDPKKVTIRIALQNAASVASLLITTECAIAEKPEPKDMMPG-GGMGGM 541
Db 480 GKIVDMFKEGIIDPLKVERIALQNAVSSLLITTEATVHEIKEEKATPAMPDMGGMGM 539
QY 542 GGMGGM 547
Db 540 GGMGGM 545

RESULT 12
US-08-466-662-6
; Sequence 6, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-466-662-6

Query Match 63.6%; Score 1723.5; DB 3; Length 546;
Best Local Similarity 64.1%; Pred. No. 3.5e-142;
Matches 350; Conservative 73; Mismatches 120; Indels 3; Gaps 2;
QY 3 SKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSGSPVITKDGVSVAKEIEL 62
Db 2 AKEIKFSDSARNLLFEGVQLHDAVKVTMGPRGNVLIQKSYGAPSIITKDGVSVAKEIEL 61
QY 63 EDKFENMGQMKVEKAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 122
Db 62 SCFVANWGAQLVKEVASKTADAAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 121
QY 123 AVAVTKELSDITKPTRDQKEIAQVGTISANSSTTTGNIIAEAMAKVGKGVITVBEAKG 182
Db 122 AAEAIINELKAKSKVGGKEEITQVATISANSDHNIKGLIADAMEKVGKGVITVBEAKG 181
QY 183 LETTLDVVEGMQDRGVLSPYFVTNPEKMCVCELDNPVILCNEKKITSMKMDLPILBOVAK 242
Db 182 IEDELVDVGMQDRGVLSPYFVTNPEKMCVCELDNPVILCNEKKITSMKMDLPILBKTWK 241
QY 243 VNRPLIIAEDVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEAIF 302
Db 242 EGKPLIIAEDIEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEAIF 301
QY 303 EBGIKLENVSLSLGTAKRVVDKENTTIIVDGAGKSEDIKARVKQIRAIQIETSSDYDR 362
Db 302 EELGLSLENAEVEFLGKAGRIVIDKNTTIIVDGKSHSDVVKDRVAQIKTQIASTTSDYDK 361
QY 363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAAVEEGIVPGGGTAFVRS 422
Db 362 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAAVEEGIVPGGGTAFVRS 421
QY 423 IKVLDDIKPADDELALAGLNIIRSLLEPLRQIAANAGYEGSIIVKVPKDFGFENAS 482
Db 422 AOKVH--LNLHDEKVGVEIIMRAIKAPLAQIAINAGYDGVVNEVEKEHGFHFNASN 479
QY 483 GEYEDLIKAGVIDPKKVTIRIALQNAASVASLLITTECAIAEKPEPKDMMPG-GGMGGM 541

Db 422 AOKVH--LNLHDEKVGVEIIMRAIKAPLAQIAINAGYDGVVNEVEKEHGFHFNASN 479
QY 483 GEYEDLIKAGVIDPKKVTIRIALQNAASVASLLITTECAIAEKPEPKDMMPG-GGMGGM 541
Db 480 GKIVDMFKEGIIDPLKVERIALQNAVSSLLITTEATVHEIKEEKATPAMPDMGGMGM 539
QY 542 GGMGGM 547
Db 540 GGMGGM 545

RESULT 13
US-08-256-847C-1
; Sequence 1, Application US/08256847C
; Patent No. 6403099
; GENERAL INFORMATION:
; APPLICANT: Rappuoli, Rino
; APPLICANT: Constantino, Paolo
; APPLICANT: No. 6403099elli, Francesco
; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
; FILE REFERENCE: CHIR-0042
; CURRENT APPLICATION NUMBER: US/08/256,847C
; CURRENT FILING DATE: 1994-11-01
; PRIOR APPLICATION NUMBER: PCT/EP93/00516
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: FI92A000058
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 546
; TYPE: PRT
; ORGANISM: H. pylori
US-08-256-847C-1

Query Match 63.6%; Score 1723.5; DB 4; Length 546;
Best Local Similarity 64.1%; Pred. No. 3.5e-142;
Matches 350; Conservative 73; Mismatches 120; Indels 3; Gaps 2;
QY 3 SKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSGSPVITKDGVSVAKEIEL 62
Db 2 AKEIKFSDSARNLLFEGVQLHDAVKVTMGPRGNVLIQKSYGAPSIITKDGVSVAKEIEL 61
QY 63 EDKFENMGQMKVEKAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 122
Db 62 SCFVANWGAQLVKEVASKTADAAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 121
QY 123 AVAVTKELSDITKPTRDQKEIAQVGTISANSSTTTGNIIAEAMAKVGKGVITVBEAKG 182
Db 122 AAEAIINELKAKSKVGGKEEITQVATISANSDHNIKGLIADAMEKVGKGVITVBEAKG 181
QY 183 LETTLDVVEGMQDRGVLSPYFVTNPEKMCVCELDNPVILCNEKKITSMKMDLPILBOVAK 242
Db 182 IEDELVDVGMQDRGVLSPYFVTNPEKMCVCELDNPVILCNEKKITSMKMDLPILBKTWK 241
QY 243 VNRPLIIAEDVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEAIF 302
Db 242 EGKPLIIAEDIEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEIDAILTGGEAIF 301
QY 303 EBGIKLENVSLSLGTAKRVVDKENTTIIVDGAGKSEDIKARVKQIRAIQIETSSDYDR 362
Db 302 EELGLSLENAEVEFLGKAGRIVIDKNTTIIVDGKSHSDVVKDRVAQIKTQIASTTSDYDK 361
QY 363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAAVEEGIVPGGGTAFVRS 422
Db 362 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAAVEEGIVPGGGTAFVRS 421
QY 423 IKVLDDIKPADDELALAGLNIIRSLLEPLRQIAANAGYEGSIIVKVPKDFGFENAS 482
Db 422 AOKVH--LNLHDEKVGVEIIMRAIKAPLAQIAINAGYDGVVNEVEKEHGFHFNASN 479
QY 483 GEYEDLIKAGVIDPKKVTIRIALQNAASVASLLITTECAIAEKPEPKDMMPG-GGMGGM 541

Db 480 GKYVDMFKEGIDPLKVERIALQNAVSVSSLLLTTEATVHEIKKEKATPAMPDMGMGM 539

QY 542 GGMGM 547

Db 540 GGMGM 545

RESULT 14

US-08-256-847C-7

; Sequence 7, Application US/08256847C

; Patent No. 6403099

; GENERAL INFORMATION:

; APPLICANT: Rappuoli, Rino

; APPLICANT: Costantino, Paolo

; APPLICANT: No. 6403099elli, Francesco

; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch

; FILE REFERENCE: CHR-0042

; CURRENT APPLICATION NUMBER: US/08/256,847C

; CURRENT FILING DATE: 1994-11-01

; PRIOR APPLICATION NUMBER: PCT/EP93/00516

; PRIOR FILING DATE: 1993-03-08

; PRIOR APPLICATION NUMBER: FI92A000058

; PRIOR FILING DATE: 1992-03-06

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7

; LENGTH: 546

; TYPE: PRT

; ORGANISM: H. pylori

; US-08-256-847C-7

Query Match 63.6%; Score 1723.5; DB 4; Length 546;

Best Local Similarity 64.1%; Pred. No. 3.5e-142;

Matches 350; Conservative 73; Mismatches 120; Indels 3; Gaps 2;

QY 3 SKEILFDKAREKLSRGVDKLANAVKVTGPKGRNVVIEKSGSPVITKGVSVAKIEL 62

Db 2 AKETKSDSARNLLFEGVRLDHAQVVTGPRGNVLQKSYGAPITKGVSVAKIEL 61

QY 63 EDKENMGAQMVKEVAPKTSIADGTTTATVLAQIYREGVKLVAGRNPMIAIKRGIDK 122

Db 62 SCPVANNGAQLVKEVASKTADAAGDGTATVLAISYFKELRNITAGANPIEVKRGMDK 121

QY 123 AVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGCVITVEAKG 182

Db 122 AAELINELKASKKVSGKEITQVATISANSDDNIGKLIADAMEKVKGVITVEAKG 181

QY 183 LETTLDVVEGKDFRGYLSFYFTNPCKMVCELDNXPYILCNKKITSMKMLPILEQVAK 242

Db 182 IEDELVDVEGQDFRGYLSFYFTVNAEKMTAQLDNAYILLTDKISSMKDILPLEKTMK 241

QY 243 VNRPLIIIAEDVEGALATLVNKLQALQVAVKATGFGERRKAMLEDITAILTGGEAIF 302

Db 242 EGKPLIIIAEDIEGEALTLLVNVKLGVLNIAAVKAFGFGDRKEMKLDAILTGQGVIS 301

QY 303 EDRIKLENVSLSLGTAKRVVIDKENTTIVDGAKSKEDIKARVKQLRAQIEETSSDYDR 362

Db 302 BELGLSLENAEVEFLGKAGRIDDKNTTIVDGKSHDDVDKRVQAQIKTQIASTTSYDK 361

QY 363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAVEEGIVPGGFATVRS 422

Db 362 EKLQERLAKLVGGVAVIKVGAASEVEKMKDRVDDALSATKAVERGIVIGGAALIRA 421

QY 423 IKVLDDTKPADDELAGNIIRSRLEPLQIAANAGYEGSIVVEKREPKDGFNAAS 482

Db 422 AQKH--LNLHDEKVGVEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKEGHEGFGNAS 479

QY 483 GEYEDLIKAGVIDPKYTRIALQNAASVASILLTTECAIAEKEPKKMPG-GGMGM 541

Db 480 GKYVDMFKEGIDPLKVERIALQNAVSVSSLLLTTEATVHEIKKEKATPAMPDMGMGM 539

QY 542 GGMGM 547

Db 540 GGMGM 545

RESULT 15

US-08-467-822-33

; Sequence 33, Application US/08467822

; Patent No. 5843460

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/08/467,822

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/447,177

; FILING DATE: 19-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/432,697

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0137-02000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 33:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 544 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-467-822-33

Query Match

Best Local Similarity 63.6%; Score 1721; DB 2; Length 544;

Matches 351; Conservative 73; Mismatches 114; Indels 4; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGPKGRNVVIEKSGSPVITKGVSVAKEL 60

Db 1 MAANKIKYNEKADARKKHGKVKTLAEAVKVTGPKGRNVVIEKSGSPVITKGVSVAKEL 60

QY 61 ELEDKENMGAQMVKEVAPKTSIADGTTTATVLAQIYREGVKLVAGRNPMIAIKRG 120

Db 61 ELEDKENMGAQMVKEVASKTADAAGDGTATVLAISYFKELRNITAGANPIEVKRG 120

QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGCVITVEEA 180

Db 121 DKAVVVVDIEKKISKEPVQHHKIAQVATISANDAEIGNLIAEAMEKVGKNGSITVEEA 180

Mon Feb 2 12:04:09 2004

181	QY	KGLETTLDVVEGMMKFDGRGYLSPFVTVNPEKMWCELDNPHYILCNPKKITSMMKMLPILQOV	240
181	Db	KGPFETVLDDVVEGMNFRGYLSSYFSTNPQETQECVLEEAALVYDKKISGDKFDLVLQOV	240
241	QY	AKVNRPLLIIAEDVEGEALATLVNKLRGALQVAVKAPGGRKKAMLEIAILITGGEA	300
241	Db	AESGRPLLIIAEDIEGALATLVNVRURAGFRVCAVAPGGRKKAMLEIAILITGSQL	300
301	QY	IFDRGKLENVLSISLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAIQIBETSSDY	360
301	Db	ISELGMKLENTTLAMLGKAKVIVSKEDTIVEGLGSKEDIESRCESIKKQIEDSTSDY	360
361	QY	DREKLQRLAKLVGGVAVIHVGATATEMKKKORVEDALNATRAAVEGIVPGGCTAFV	420
361	Db	DPEKLQRLAKLVGGVAVIRVGAATEIEMKKORVDQAQATLAAVEEGILPGGCTALV	420
421	QY	RSIKVLVDDIKP--ADDDDELAGLNIIRSLPELQIAANAGYEGSVVEKVPKDGFGF	478
421	Db	RCPTPLFAFPIILTNEDEQIGARVLKALSAPLKQIAANAGKEGAILICQVLSRSSSEGY	480
479	QY	NAASAGEYEDLIKAGVIDPKVKTRIALQNAASVASLLLTTECAIAEKPEPKKDM--PMGG	536
481	Db	DALRDAYTDMIAGILDPKVKTRCALBSAASVAGILLTTEALADIPEKSSSAPMPGA	540
537	QY	GM 538	
541	Db	GM 542	

Search completed: January 28, 2004, 13:18:23
Job time : 24 secs

1	1862.5	68.8	547	12	US-10-369-493-17524	Sequence 17524, A
2	1857	68.6	547	12	US-10-369-493-10014	Sequence 10014, A
3	1840.5	68.0	544	12	US-10-369-493-11760	Sequence 11760, A
4	1839.5	67.9	542	12	US-10-369-493-7917	Sequence 7917, Ap
5	1835.5	67.8	547	12	US-10-369-493-10295	Sequence 10295, A
6	1834	67.7	541	12	US-10-369-493-9287	Sequence 9287, Ap
7	1833	67.7	544	12	US-10-369-493-15468	Sequence 15468, A
8	1833	67.7	544	12	US-10-369-493-16217	Sequence 16217, A
9	1828.5	67.5	542	12	US-10-369-493-15265	Sequence 15265, A
10	1826	67.4	542	12	US-10-369-493-11922	Sequence 11922, A
11	1812.5	66.9	551	12	US-10-369-493-12330	Sequence 12330, A
12	1811	66.9	540	12	US-10-369-493-15938	Sequence 15938, A
13	1810	66.8	552	12	US-10-369-493-12170	Sequence 12170, A
14	1786	66.0	545	12	US-10-369-493-167	Sequence 167, App
15	1780.5	65.7	544	12	US-10-369-493-17132	Sequence 17132, A

[illegible]

17

[illegible]

Db 121 DKAVIAAVTELKISKPTSDDKAIAQVATISANSDESIGNIIAEAMKVKGEVITIBEG 180
 Qy 181 KGETTLDVVEGMDRGLSPYFVTNPEKVCWCELDNPIYLCNEKKTSMKMDLPILEOV 240
 Db 181 TTLENDLVVEGMDRGLSPYFVTNPEKVCWCELDNPIYLCNEKKTSMKMDLPILEOV 240
 Qy 241 AKVNEPLIIADVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 Db 241 AKVNEPLIIADVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 Qy 301 IPEDRGIKLENVSLSSLTAKRVVIDKENTTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
 Db 301 ISEEVGLSLFKATTHGKAKVRVSKENTTTIIDGIDNDALNGRVKQIKTQIETSSDY 360
 Qy 361 DREKLOERLAKLVGVAVIHVCAATETEMKEKDRVEDALNATRAAVEGIVPGGTAFFV 420
 Db 361 DREKLOERLAKLVGVAVIHVCAATETEMKEKDRVEDALNATRAAVEGIVPGGTAFFV 420
 Qy 421 RSKVLDDIKPADDDELAGLNIIRSLPELRQIAANAGYEGSVVVEKVRPEPKD-GFGF 480
 Db 421 RAITAISNLKGANEDOTHGIIQALRAMEAPLEIRIVANAGEEPPSVILNKVKEGKDNFGYNA 480
 Qy 481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASILLTTECAIAEKPEPKDMPMP---GGG 537
 Db 481 ATGEPGDMVNLGILDPKTVTRGALQNAASVAGLMIITTEAMVAE--APKKDEPTTPAAGG 538
 Qy 538 MCGMGMD 545
 Db 539 MCGMGMD 546

RESULT 2

US-10-369-493-10014
 ; Sequence 10014, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 10014
 ; LENGTH: 547
 ; TYPE: PRT
 ; ORGANISM: magnetite-containing magnetic coccus
 US-10-369-493-10014

Query Match 68.6%; Score 1857; DB 12; Length 547;
 Best Local Similarity 69.5%; Pred. No. 3.4e-138;
 Matches 382; Conservative 72; Mismatches 90; Indels 6; Gaps 4;
 Qy 1 MASKEILFIDAKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSGFSPVITKDGVSVAKEI 60
 Db 1 MAKEVKGFEARAKMLNGVILANAVKVTLPKGRNVLDKSWGAPMTKDGVSVAKEI 60
 Qy 61 ELEDKFNMGAAQWKEVAPKTSIADGTTTATVLAQAIYREGVLAAGRNPMIAIKRGI 120
 Db 61 ELEDKFNMGAAQWKEVAPKTSIADGTTTATVLAQAIYREGVLAAGRNPMIAIKRGI 120
 Qy 121 DKAVAVTKELSDITKTRQKETAQVGTISANSDDTIGNIIAEAMKVKGGVITVEEA 180
 Db 121 DLAVEAVVGLKISREAVNSOEIAQVGAISANSKDVGVGMIAEMDKVKEGVITVEEA 180
 Qy 181 KGETTLDVVEGMDRGLSPYFVTNPEKVCWCELDNPIYLCNEKKTSMKMDLPILEOV 240

Db 181 KGETTLDVVEGMDRGLSPYFVTNADKMLVQMENPILILVEKKISNLOQLILEGA 240
 Qy 241 AKVNEPLIIADVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 Db 241 VOSSRPLIIADVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 Qy 301 IPEDRGIKLENVSLSSLTAKRVVIDKENTTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
 Db 301 VSEDVGKLENVMTDMLGMAKSIIVTKEDTTIIDGAGDHEAIKARVNOIRAOIETSSDY 360
 Qy 361 DREKLOERLAKLVGVAVIHVCAATETEMKEKDRVEDALNATRAAVEGIVPGGTAFFV 420
 Db 361 DREKLOERLAKLVGVAVIKVGGAATEVEKERRDVRDDALHATRAAVEGIVPGGVALI 420
 Qy 421 RSKVLDDIKPADDDELAGLNIIRSLPELRQIAANAGYEGSVVVEKVRPEPKD-GFGF 478
 Db 421 RAREASLTNLQANHDQVGINIVFALPELRITASNAGESVVVNVVETKETNFGF 480
 Qy 479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASILLTTECAIAEKPEPKDMP-MPGGG 537
 Db 481 NAATGVYEDLVASGVIDPAKVVRHALQAAASVAGLMIITTEAMVAELP---KDEPAMPGG 537
 Qy 538 MCGMGMD 547
 Db 538 MCGMGMD 547

RESULT 3

US-10-369-493-11760
 ; Sequence 11760, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 11760
 ; LENGTH: 544
 ; TYPE: PRT
 ; ORGANISM: Agrobacterium tumefaciens
 US-10-369-493-11760

Query Match 68.0%; Score 1840.5; DB 12; Length 544;
 Best Local Similarity 69.2%; Pred. No. 6.9e-137;
 Matches 376; Conservative 69; Mismatches 95; Indels 3; Gaps 3;
 Qy 1 MASKEILFIDAKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSGFSPVITKDGVSVAKEI 60
 Db 1 MAKEVKGFEARSAREKMLKGVLDLADAVKVTLPKGRNVLDKSGFAPRITKDGVSVAKEI 60
 Qy 61 ELEDKFNMGAAQWKEVAPKTSIADGTTTATVLAQAIYREGVLAAGRNPMIAIKRGI 120
 Db 61 ELEDKFNMGAAQWKEVAPKTSIADGTTTATVLAQAIYREGVLAAGRNPMIAIKRGI 120
 Qy 121 DKAVAVTKELSDITKTRQKETAQVGTISANSDDTIGNIIAEAMKVKGGVITVEEA 180
 Db 121 DLAVEAVVGLKISREAVNSOEIAQVGAISANSKDVGVGMIAEMDKVKEGVITVEEA 180
 Qy 181 KGETTLDVVEGMDRGLSPYFVTNPEKVCWCELDNPIYLCNEKKTSMKMDLPILEOV 240
 Db 181 KTAETELEVVEGMDRGLSPYFVTNPEKVCWCELDNPIYLCNEKKTSMKMDLPILEOV 240
 Qy 241 AKVNEPLIIADVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAITLTGGEA 300

Db 241 VQTGKPLVIAEDVEGEALATLVNKLRLGGLKIAAVKAPGFGDRRKAMLEDIAILLTGTV 300
 Qy 301 IFEDRGKLENVSLSSGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360
 Db 301 ISEDIGIKLESVTLDMGLCKSKVSISKENTTIVDGAGKSDIEGRVAQIKAOIETSSDY 360
 Qy 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAABEGIVPGGTAFFV 420
 Db 361 DREKLOERLAKLVGGVAVIRVGGSTEVEVKEDKORIDDALNATRAAABEGIVPGGTAFFV 420
 Qy 421 RSKVLDDIKPADDELALNLIIRSLPEERQIAANAGYEGSIIVVEKVRP-KDGFQFN 479
 Db 421 RSSTKI-TVKGNDDEAGINIRKALQSLVRQIAENAGDEASIVVGGKILDKNEDNYGN 479
 Qy 480 AASGEYEDLIKAGVIDPKKVTTRIALQNAASVASLLLTTECAIAEKPEPKDMP-MPGGGM 538
 Db 480 AOTGEYGDIALGIVDPKVVTRTALQNAASVASLLLTTECAIAELPKESAMPQMPGGM 539
 Qy 539 GGM 541
 Db 540 GGM 542

RESULT 4

US-10-369-493-7917
 ; Sequence 7917, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 7917
 ; LENGTH: 547
 ; TYPE: PRT
 ; ORGANISM: Rhodobacter sphaeroides
 US-10-369-493-7917

Query Match 67.9%; Score 1839.5; DB 12; Length 547;
 Best Local Similarity 67.9%; Pred. No. 8.3e-137;
 Matches 372; Conservative 70; Mismatches 103; Indels 3; Gaps 2;
 Qy 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
 Db 1 MAADKVKFTDARDMLRGVNLIDAVKVTGLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
 Qy 61 ELEDKFNMGQVMKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMALKRG 120
 Db 61 ELSDKFNMGQVMKEVASTNDEAGDGTATVLAQAIYREGVKLVAAAGRNPMALKRG 120
 Qy 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGKGVITVEEA 180
 Db 121 DLATSKVWEIAKAAARPVNDSEHVAQVGTISANGEAQIGRFIADAMQKVGNEGVITVEEN 180
 Qy 181 KGLTEVVEVGGMDFRGVLSFYFVTNPKEKVCMLDNPYLCKEIKITSKMLPILQV 240
 Db 181 KGLTEVVEVGGMDFRGVLSFYFVTNPKEKVCMLDNPYLCKEIKITSKMLPILQV 240
 Qy 241 AKVNRPLIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGDRRKAMLEDIAILLTGGA 300
 Db 241 IQSQKPLIIAEDVEGEALATLVNKLRLGKIAAIVKAPGFGDRRKAMLEDIAILLTGGA 300
 Qy 301 IFEDRGKLENVSLSSGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 360

Db 301 ISEDLGMKLENTIDMLGRAKKISINKNTTIVDNGDKAEIDARVAQIRNOIETSSDY 360
 Qy 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAABEGIVPGGTAFFV 420
 Db 361 DREKLOERLAKLVGGVAVIRVGGSTEVEVKEDKORIDDALNATRAAABEGIVPGGTAFFV 420
 Qy 421 RSKVLDDIKPADDELALNLIIRSLPEERQIAANAGYEGSIIVVEKVRP-KDGFQFN 479
 Db 421 QGGKALDGLTGENPDQAGITIVRRALQSLVRQIAENAGDEASIVVGGKILDKNEDNYGN 480
 Qy 480 AASGEYEDLIKAGVIDPKKVTTRIALQNAASVASLLLTTECAIAEKPEPKDMP-MPGGGM 539
 Db 481 AOTGEYGDIMFKEGVIDPAKVVTRTALQNAASVASLLLTTECAIAELPKESAMPQMPGGM 538
 Qy 540 GGM 547
 Db 539 GGM 546

RESULT 5

US-10-369-493-10295
 ; Sequence 10295, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 10295
 ; LENGTH: 542
 ; TYPE: PRT
 ; ORGANISM: Cytophaga hutchinsonii
 US-10-369-493-10295

Query Match 67.8%; Score 1835.5; DB 12; Length 542;
 Best Local Similarity 67.6%; Pred. No. 1.7e-136;
 Matches 365; Conservative 77; Mismatches 97; Indels 1; Gaps 1;
 Qy 3 SKEILFDKAREKLSRGVDKLANAVKVTGLGPKGRNVVIEKSPGVITKDGVSVAKEI 62
 Db 2 AKQILFDKAREKLSRGVDKLANAVKVTGLGPKGRNVVIEKSPGVITKDGVSVAKEI 61
 Qy 63 EDKFNMGQVMKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMALKRG 122
 Db 62 KDAIENMGQVMKEVASKVATADQAGDGTATVLAQAIYREGVKLVAAAGRNPMALKRG 121
 Qy 123 AVAVATKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGKGVITVEEA 182
 Db 122 AVSAIIADLVKQSKISNSNEIAQVATISANNDEHIGKMLAHAMDKVKGKGVITVEEA 181
 Qy 183 LETTLDVVEGKDFRGVLSFYFVTNPKEKVCMLDNPYLCKEIKITSKMLPILQV 242
 Db 182 TETEVTVEGMDFRGVLSPFYFVTNPKEKVCMLDNPYLCKEIKITSKMLPILQV 241
 Qy 243 VNRPLIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGDRRKAMLEDIAILLTGGA 302
 Db 242 SGRPLIIAEDVDGEALATLVNKLRLGALQVAVKAPGFGDRRKAMLEDIAILLTGGA 301
 Qy 303 EDRGKLENVSLSSGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAOIETSSDY 362
 Db 302 EESGYKLENATLEYLGTAEKINIDKNTTIVVNGSLKONIVARVNOIKSQMENTTSSDY 361
 Qy 363 EKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAABEGIVPGGTAFFV 422

Db 362 EKLOERLAKSCGVAIMVIGATVEVMEKKORVDLDALHATRAAEEGIVPGGGVALIRA 421
Qy 423 IKVLDIKPADDDDELAGNIIRSRLEPLROIAANAGYEGSVIVKVEPKDGFNFNAAS 482
Db 422 GAALDNVAFHNEDEKGTQIIRKATATESPLRAIVNAGLEGSVIVQVKEGTGDYGNARE 481
Qy 483 GEYEDLIKAGVDPKPVTRIALONAAASVALLTTCECAIAEKPEPKDM-PMPGGGGMGM 541
Db 482 DRYEAMIAAGIIDPTKVTRLAENAAASVALLTTCEVVADEPEKKSAMPMPGGGGMGM 541

RESULT 6
US-10-369-493-9287
; Sequence 9287, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9287
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9287

Query Match 67.7%; Score 1834; DB 12; Length 541;
Best Local Similarity 66.3%; Pred. No. 2.2e-136;
Matches 360; Conservative 80; Mismatches 101; Indels 2; Gaps 1;

Qy 1 MASKEILFPAKAREKLSRGVDKLANAVKVTLPKGRNVVLEKSFSGSVITKDGVSVAKEI 60
Db 1 MAKEILFPAKAREKLSRGVDKLANAVKVTLPKGRNVVLEKSFSGSVITKDGVSVAKEI 60
Qy 61 ELEDKFNENGAQVMKEVAPKTSIADGTTTAVLAQAIYREGVKLVAAENPMALKRG 120
Db 61 ELADKFNENGAQVMKEVAPKTSIADGTTTAVLAQAIYREGVKLVAAENPMALKRG 120
Qy 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEEA 180
Db 121 DKAVIAAVTELKKSITSDDKALAQVATISANSDESIGNIIAEAMAKVGKGVITVEEG 180
Qy 181 KGLTTLDDVVEGKMPDRGYLSFYFTNPEKMWCELDNPNYILCNEKKTSMKMLPILQV 240
Db 181 TTELENDLVVEGKMPDRGYSSPYFINNQSQIIVELDNPNYILCNEKKTSMKMLPILQV 240
Qy 241 AKVNRPLLIITAEVDEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGEEA 300
Db 241 AKESKPLLIIVAEVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGEEA 300
Qy 301 IFEDRGIKLENVSLSSLTGAKRVVVDKENTTIIDGAGSDIARVQIRAKIIEETSSDY 360
Db 301 ISEVGLSLEKATTSHTSHGAKRVVSKENTTIIDGAGSDIARVQIRAKIIEETSSDY 360
Qy 361 DREKLOERLAKLVGGVAIVHGAATEMEKEDKORVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLOERLAKLVGGVAIVHGAATEMEKEDKORVEDALNATRAAEEGIVPGGGTAFV 420
Qy 421 RSIKVLDIKPADDDDELAGNIIRSRLEPLROIAANAGYEGSVIVKVEPKDGFNFNA 480
Db 421 RAITAIENLKGANEDQTHGIIALRAEAPLREIVANAGEPSVILNKVKEGKNFNGYNA 480
Qy 481 ASGEYEDLIKAGVIDPKKVTTRIALONAAASVALLTTCECAIAEKPEPKDM-PMPGGGGMGM 540
Db 481 ASGEYEDLIKAGVIDPKKVTTRIALONAAASVALLTTCECAIAEKPEPKDM-PMPGGGGMGM 540

Db 481 ATGFBGDMVKLGILDPKVTKARTALONAAASIAGLMITTEAMVAE--APKDEBTPPAAGG 538
Qy 541 MGG 543
Db 539 MGG 541

RESULT 7
US-10-369-493-15468
; Sequence 15468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15468
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15468

Query Match 67.7%; Score 1833; DB 12; Length 544;
Best Local Similarity 66.8%; Pred. No. 2.7e-136;
Matches 365; Conservative 82; Mismatches 95; Indels 4; Gaps 3;

Qy 1 MASKEILFPAKAREKLSRGVDKLANAVKVTLPKGRNVVLEKSFSGSVITKDGVSVAKEI 60
Db 1 MAADKIRFGEADARTVMRGVNLANAVKATLGPGRNVVLEKSFSGSVITKDGVSVAKEI 60
Qy 61 ELEDKFNENGAQVMKEVAPKTSIADGTTTAVLAQAIYREGVKLVAAENPMALKRG 120
Db 61 ELADKFNENGAQVMKEVAPKTSIADGTTTAVLAQAIYREGVKLVAAENPMALKRG 120
Qy 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEEA 180
Db 121 DQAVKAAVIELKNSIPTDDKALAQVGTISANSDESIGNIIAEAMAKVGKGVITVEEG 180
Qy 181 KGLTTLDDVVEGKMPDRGYLSFYFTNPEKMWCELDNPNYILCNEKKTSMKMLPILQV 240
Db 181 SGLLENLDVVEGKMPDRGYLSFYFINNQSQSADLDDPFILLHDKKLSNVRLDLPVLEGV 240
Qy 241 AKVNRPLLIITAEVDEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGEEA 300
Db 241 AKAGKPLLIIVAEVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGEEA 300
Qy 301 IFEDRGIKLENVSLSSLTGAKRVVVDKENTTIIDGAGSDIARVQIRAKIIEETSSDY 360
Db 301 ISEVGLSLEKATTSHTSHGAKRVVSKENTTIIDGAGSDIARVQIRAKIIEETSSDY 360
Qy 361 DREKLOERLAKLVGGVAIVHGAATEMEKEDKORVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLOERLAKLVGGVAIVHGAATEMEKEDKORVEDALNATRAAEEGIVPGGGTAFV 420
Qy 421 RSIKVLDIKPADDDDELAGNIIRSRLEPLROIAANAGYEGSVIVKVEPKDGFNFNA 480
Db 421 RALVAVGNLKGANEDQTHGIIALRAEAPLREIVANAGEPSVILNKVKEGKNFNGYNA 480
Qy 481 ASGEYEDLIKAGVIDPKKVTTRIALONAAASVALLTTCECAIAEKPEPKDM-PMPGGGGMGM 538
Db 481 ANGEFGDMVEGILDPTKVTTRIALONAAASVALLTTCECAIAEKPEPKDM-PMPGGGGMGM 538
Qy 539 GGMGM 544
Db 539 GGMGM 544

Db 539 GGMGM 544

RESULT 8

US-10-369-493-16217
; Sequence 16217, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 16217

; LENGTH: 544

; TYPE: PRT

; ORGANISM: Xanthomonas campestris

US-10-369-493-16217

Query Match 67.7%; Score 1833; DB 12; Length 544;

Best Local Similarity 66.8%; Pred. No. 2.7e-136;

Matches 365; Conservative 82; Mismatches 95; Indels 4; Gaps 3;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSGSPVITKDGVSVAKEI 60

Db 1 MAAKDIRFGEDARTMRVGRVNVLANAVKATLGPGRNVVIEKSGCAPITKDGVSVAKEI 60

QY 61 ELEDKFENMGAMQVKEVAPKTSIDTAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120

Db 61 ELADKFENMGAMQVKEVARSRTNDNAGDGTATVLAQALIREGAKAVAAAGNPNMDLKRGI 120

QY 121 DKAVAVTKELSDITKTRDQKETAQVGTISANSDDTIGNIIAEMAKVKGGVITVEEA 180

Db 121 DQAVKAAVIELKNISKFTDDKALAQVGTISANSDESIGNIIAEMAKVKGGVITVEEG 180

QY 181 KGLFTLLDVVEGKFDGYSLPYFVTNPKEKWCCLDNPYILCNEKKTITSMKMLPILEQV 240

Db 181 SGLNELDVVEGQFDRGYSLPYFINNQSQSADLDDPFLLHDKKLSNVRLDLPVLEGV 240

QY 241 AKVNRPLIIIAEDVEGEALATLVVKNLKGALQVAVKAPGFGERRKAMLEDAIALLTGEEA 300

Db 241 AKAGKPLIIIAEDVEGEALATLVVNTIRGIVKVVAVKAPGFGDRRRKAMLEDAVLTGTV 300

QY 301 IFEDRGIKLENVSLSSIGTAKRVVIDKENTTIIVDGAGKSEDIKARVKQIRAOIEETSSDY 360

Db 301 ISEVGLALEKATIKOLGRKVVQVSKENTTIIDGAGDSATIEARVGQIKQIEDTSSDY 360

QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALNATRAAVEGIVPGGGTAFV 420

Db 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALHATRAAVEGIVPGGGVALV 420

QY 421 RSLKVLDDIKPADDDDELAGNIIRSLRLEPLRQIAANAGYEGSVIVVEKVRREP-KDGFEN 480

Db 421 RALVAVGNITGANEDQTHGIQIALRAMEAPLREIVANAGEPSVLNKKVKEGTNGVYNA 480

QY 481 ASGEYEDLIKAGVIDPKKVTIRIALQNAASVASLLLTTECAIAEKPEPKOMP-MP-GGGM 538

Db 481 ANGEFGDMVEFGILDFTKVRSLQNAASITAGLMTITTEAMVAD--APKKDEPAMPAGGM 538

QY 539 GGMGM 544

Db 539 GGMGM 544

RESULT 9

RESULT 10

US-10-369-493-11922

; Sequence 11922, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

US-10-369-493-15265

; Sequence 15265, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 15265

; LENGTH: 543

; TYPE: PRT

; ORGANISM: Agrobacterium tumefaciens

US-10-369-493-15265

Query Match 67.5%; Score 1828.5; DB 12; Length 543;

Best Local Similarity 68.9%; Pred. No. 6.1e-136;

Matches 374; Conservative 70; Mismatches 96; Indels 3; Gaps 3;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSGSPVITKDGVSVAKEI 60

Db 1 MAAEKVPKASAREKMLKGVLDLADAVKTLGPKGRNVVIEKSGCAPITKDGVSVAKEI 60

QY 61 ELEDKFENMGAMQVKEVAPKTSIDTAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120

Db 61 ELEDKFENMGAMQVKEVARSRTNDNAGDGTATVLAQALIREGAKAVAAAGNPNMDLKRGI 120

QY 121 DKAVAVTKELSDITKTRDQKETAQVGTISANSDDTIGNIIAEMAKVKGGVITVEEA 180

Db 121 DLVAEVVVKDLQAKAKKINTSEEAQVGTISANGERTQIGLDIAEAMQVNEGVITVEEA 180

QY 181 KGLFTLLDVVEGKFDGYSLPYFVTNPKEKWCCLDNPYILCNEKKTITSMKMLPILEQV 240

Db 181 KTAETELEVVEGQFDRGYSLPYFVTNPKEKWCCLDNPYILCNEKKTITSMKMLPILEAV 240

QY 241 AKVNRPLIIIAEDVEGEALATLVVKNLKGALQVAVKAPGFGERRKAMLEDAIALLTGEEA 300

Db 241 VQTKPLIIIAEDVEGERLATLVVKNLKGALQVAVKAPGFGDRRRKAMLEDAIALLTGTV 300

QY 301 IFEDRGIKLENVSLSSIGTAKRVVIDKENTTIIVDGAGKSEDIKARVKQIRAOIEETSSDY 360

Db 301 ISEDGKLESVTLDMIGSKSKVSIKENTTIIVDGAGQKSDIEGRVAQIRAOIEETSSDY 360

QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALNATRAAVEGIVPGGGTAFV 420

Db 361 DREKLOERLAKLVGGVAVIRVGGSTEVEVEKDRIDDLNATRAAVEGIVPGGGVALL 420

QY 421 RSLKVLDDIKPADDDDELAGNIIRSLRLEPLRQIAANAGYEGSVIVVEKVRREP-KDGFEN 479

Db 421 RSSTKI-TVKGVNDDQEGAGINIVRKALQSLVRQIAENAGDEASIVVVKILDKNEDNYGN 479

QY 480 AASGEYEDLIKAGVIDPKKVTIRIALQNAASVASLLLTTECAIAEKPEPKOMP-MPGGGM 538

Db 480 AQTGEYEDLIKAGVIDPKKVTIRIALQNAASVASVLTITTEAMIAELPKESAMPQMGGM 539

QY 539 GGM 541

Db 540 GGM 542

Mon Feb 2 12:04:09 2004

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US 60/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 11922
LENGTH: 542
TYPE: PRT
ORGANISM: Mesorhizobium loti
US-10-369-493-11922

Query Match 67.4%; Score 1826; DB 12; Length 542;
Best Local Similarity 67.8%; Pred. No. 9.5e-136; Indels 2; Gaps 2;
Matches 366; Conservative 80; Mismatches 92

QY 1 MASKEILPAKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSPVITKDGVSVAKEI 60
DB 1 MAKEVKFHSRDAREKRLGVNLAADVAVKVTLPKGRNVVIEKSPVITKDGVSVAKEI 60
QY 61 ELEDKFENNGAQMVKVEAPKTSIDAGDGTITATVLAQAIYREGVKLVAAGRNPMATKRG 120
DB 61 ELEDKFENNGAQMVKVEAPKTSIDAGDGTITATVLAQAIYREGVKLVAAGRNPMATKRG 120
QY 121 DKAVAVATKELSDITKPTDRDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEA 180
DB 121 DKAVAVATKELSDITKPTDRDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEA 180
QY 181 KGLTLDVVEGKMFDRGYLSPFYVTNPEKVMCELDNPYLKNEKITSKMDLPILEQV 240
DB 181 KGLTLDVVEGKMFDRGYLSPFYVTNPEKVMCELDNPYLKNEKITSKMDLPILEQV 240
QY 181 KTAETELVVEGKMFDRGYLSPFYVTNPEKVMCELDNPYLKNEKITSKMDLPILEQV 240
DB 181 KTAETELVVEGKMFDRGYLSPFYVTNPEKVMCELDNPYLKNEKITSKMDLPILEQV 240
QY 241 AKVNRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEDAIILTGGEA 300
DB 241 AKVNRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEDAIILTGGEA 300
QY 301 IFEDRGKLENVSLSGTAKRVVIDKENTTIIVDGAGKSEDIKARVKQIRAOIEETSSDY 360
DB 301 IFEDRGKLENVSLSGTAKRVVIDKENTTIIVDGAGKSEDIKARVKQIRAOIEETSSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALNATRAAEEGIVFGGGTAFV 420
DB 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALNATRAAEEGIVFGGGTAFV 420
QY 421 RSIKVLDDIKP--ADDELAGLNIIRSLPEERQIATAANAGYEGSVIVKVRPKDG-FG 477
DB 421 RSIKVLDDIKP--ADDELAGLNIIRSLPEERQIATAANAGYEGSVIVKVRPKDG-FG 477
QY 478 FNAASCEYEDLIKAGVIDPKVTRIALQNAASVASILLTTECAIAEKPEPK--DMP-- 532
DB 478 FNAASCEYEDLIKAGVIDPKVTRIALQNAASVASILLTTECAIAEKPEPK--DMP-- 532
QY 533 MPGGGGMGGMGMD 545
DB 533 MPGGGGMGGMGMD 545

RESULT 11
US-10-369-493-12330
Sequence 12330, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12330
LENGTH: 551
TYPE: PRT
ORGANISM: Mesorhizobium loti
US-10-369-493-12330

Query Match 66.9%; Score 1812.5; DB 12; Length 551;
Best Local Similarity 67.8%; Pred. No. 1.1e-134; Indels 11; Gaps 5;
Matches 375; Conservative 71; Mismatches 96

QY 1 MASKEILPAKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSPVITKDGVSVAKEI 60
DB 1 MAKEVKFHSRDAREKRLGVNLAADVAVKVTLPKGRNVVIEKSPVITKDGVSVAKEI 60
QY 61 ELEDKFENNGAQMVKVEAPKTSIDAGDGTITATVLAQAIYREGVKLVAAGRNPMATKRG 120
DB 61 ELEDKFENNGAQMVKVEAPKTSIDAGDGTITATVLAQAIYREGVKLVAAGRNPMATKRG 120
QY 121 DKAVAVATKELSDITKPTDRDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEA 180
DB 121 DKAVAVATKELSDITKPTDRDQKEIAQVGTISANSDDTIGNIIAEMAKVKGKGVITVEEA 180
QY 181 KGLTLDVVEGKMFDRGYLSPFYVTNPEKVMCELDNPYLKNEKITSKMDLPILEQV 240
DB 181 KGLTLDVVEGKMFDRGYLSPFYVTNPEKVMCELDNPYLKNEKITSKMDLPILEQV 240
QY 241 AKVNRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEDAIILTGGEA 300
DB 241 AKVNRPLLIADVEGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEDAIILTGGEA 300
QY 301 IFEDRGKLENVSLSGTAKRVVIDKENTTIIVDGAGKSEDIKARVKQIRAOIEETSSDY 360
DB 301 IFEDRGKLENVSLSGTAKRVVIDKENTTIIVDGAGKSEDIKARVKQIRAOIEETSSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALNATRAAEEGIVFGGGTAFV 420
DB 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKDRVEDALNATRAAEEGIVFGGGTAFV 420
QY 421 RSIKVLDDIKP--ADDELAGLNIIRSLPEERQIATAANAGYEGSVIVKVRPKDG-FG 477
DB 421 RSIKVLDDIKP--ADDELAGLNIIRSLPEERQIATAANAGYEGSVIVKVRPKDG-FG 477
QY 478 FNAASCEYEDLIKAGVIDPKVTRIALQNAASVASILLTTECAIAEKPEPK--DMP-- 532
DB 478 FNAASCEYEDLIKAGVIDPKVTRIALQNAASVASILLTTECAIAEKPEPK--DMP-- 532
QY 533 MPGGGGMGGMGMD 545
DB 533 MPGGGGMGGMGMD 545

RESULT 12
US-10-369-493-15838
Sequence 15838, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15838
 ; LENGTH: 540
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15838

Query Match 66.9%; Score 1811; DB 12; Length 540;
 Best Local Similarity 66.6%; Pred. No. 1.5e-134;
 Matches 361; Conservative 82; Mismatches 95; Indels 4; Gaps 3;

QY 1 MASKEIILFDKARBKLSRGVDKLANAVKVTLPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 Db 1 MAARKDIRFGEDARTRMVRGVNVLANAVKATLPGKGRNVVIEKSPGAPITITKDGVSVAKEI 60
 QY 61 ELEKPFENMGQAMQVEKAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 Db 61 ELADKPFENMGQAMQVEKASRTNDNAGDGTATVLAQALIREGAKAVAAAGNPNMDLKRGI 120
 QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAAMAKVKGKGVITVEEA 180
 Db 121 DQAVKAAVIELKNISKPTTDDKATAQVGTISANSDESIGNIIAAMAKVKGKGVITVEEG 180
 QY 181 KLETTLDVVEGKMFDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKMDLPILFQV 240
 Db 181 SGLNELDVVEGKMFDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKMDLPILFQV 240
 QY 241 AKVNRPLIIIAEDVEGALATLVVKNLKGALQVAVKAPGGERKKALEDIAIITGGEA 300
 Db 241 AKAGKPLIIIAEDVEGALATLVVNTIRGIVKVVAVKAPGGERKKALEDIAIITGGEA 300
 QY 301 IFEDRGKLENVSLSSIGTAKRVVIDKENTTIVDAGKSEDIKARVKQIRAIQIEETSSDY 360
 Db 301 ISEVGLALEKATIKOLGRKVVQVSKENTTIVDAGSATSIEARVQIKQIEETSSDY 360
 QY 361 DREKLOERLAKLVGGVAIVHGAATEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
 Db 361 DREKLOERLAKLVGGVAIVKVGASTEEMKEKARVEDALHATRAAEEGIVPGGGTAFV 420
 QY 421 RSIVKLVDDIKPADDDELALGNIIRSELEPLRQIAANAGVEGSIIVKVEPKDGPFGFNA 480
 Db 421 RALVAVNGLTGANEEDQTHGTOIALRAMEAPLREIVANAGEEPSVILNKVKEGTGNYGNA 480
 QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASILLITTECAIAEKPEPKDMP-MP-CGGM 538
 Db 481 ANGEFGDMVEFGILDPKTVTRIALQNAASVAGLLVITTEAMVAD--APKDEPAMPAGGM 538
 QY 539 GG 540
 Db 539 GG 540

RESULT 13

US-10-369-493-12170
 ; Sequence 12170, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 12170
 ; LENGTH: 552
 ; TYPE: PRT

; ORGANISM: Mesorhizobium loti
 US-10-369-493-12170

Query Match 66.8%; Score 1810; DB 12; Length 552;
 Best Local Similarity 67.6%; Pred. No. 1.8e-134;
 Matches 373; Conservative 72; Mismatches 99; Indels 8; Gaps 4;

QY 1 MASKEIILFDKARBKLSRGVDKLANAVKVTLPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 Db 1 MAARKDIRFGEDARTRMVRGVNVLANAVKATLPGKGRNVVIEKSPGAPITITKDGVSVAKEI 60
 QY 61 ELEKPFENMGQAMQVEKAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 Db 61 ELEKPFENMGQAMQVEKASRTNDNAGDGTATVLAQALIREGAKAVAAAGNPNMDLKRGI 120
 QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAAMAKVKGKGVITVEEA 180
 Db 121 DLAVSDVVVTLIKNATKIKTSEEVAQVGTIAGNGDESIVGKMAEAMQKVGNEGVITVEEA 180
 QY 181 KLETTLDVVEGKMFDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKMDLPILFQV 240
 Db 181 KTAETELVVEGKMFDRGYLSPYFVTNADKMAVLEDAIILLHEKKLSNLQAMLPLVEAV 240
 QY 241 AKVNRPLIIIAEDVEGALATLVVKNLKGALQVAVKAPGGERKKALEDIAIITGGEA 300
 Db 241 VQTSKPLIIIAEDVEGALATLVVKNLKGALQVAVKAPGGERKKALEDIAIITGGEA 300
 QY 301 IFEDRGKLENVSLSSIGTAKRVVIDKENTTIVDAGKSEDIKARVKQIRAIQIEETSSDY 360
 Db 301 ISEDLGKLENVSLSSIGTAKRVVIDKENTTIVDAGKSEDIKARVKQIRAIQIEETSSDY 360
 QY 361 DREKLOERLAKLVGGVAIVHGAATEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
 Db 361 DREKLOERLAKLVGGVAIVRGGATEVEVEKEDRVEDALNATRAAEEGIVPGGGTAFV 420
 QY 421 RSIVKLVDDIKPADDDELALGNIIRSELEPLRQIAANAGVEGSIIVKVEPKDGPFGFNA 479
 Db 421 RASLSINAV--GANSDDTAGISIVRRALQAPARQIAANAGAEASIVAKILLENKGTATGFN 479
 QY 480 ASGEYEDLIKAGVIDPKVTRIALQNAASVASILLITTECAIAEKPEPKK---DMP--M 533
 Db 480 AQTGEYEDLIKAGVIDPKVTRIALQNAASVASILLITTECAIAEKPEPKK---DMP--M 539
 QY 534 PGGMGMGMGM 545
 Db 540 GGGGGMGMGM 551

RESULT 14

US-10-369-493-167
 ; Sequence 167, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 167
 ; LENGTH: 545
 ; TYPE: PRT

; ORGANISM: Aquifex aeolicus
 US-10-369-493-167

Query Match

66.0%; Score 1786; DB 12; Length 545;

Best Local Similarity 64.9%; Pred. No. 1.4e-132;
Matches 352; Conservative 94; Mismatches 86; Indels 10; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGKGRNVLEKFGSPVITKDGVSVAKEI 60
DB 1 MAARAILYNEBARAKLKGVDKLANAVKVTGLGKGRNVLEKFGSPVITKDGVSVAKEI 60

QY 61 ELEDKFNMGAMQVEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
DB 61 ELKDKFENIGAGLVKEVASKTADVAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRG 120

QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAABAMAKVGKGVITVEEAKG 180
DB 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAABAMAKVGKGVITVEEAKG 180

QY 181 KGLTTLDDVVEGKMPDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKDMPLILEQV 240
DB 181 KSAETTLVVGKMQFDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKDMPLILEQV 240

QY 241 AKVNPILLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLGGEA 300
DB 241 VESGPELLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLGGEA 300

QY 301 IEDRGIKLENVSLSSLTGAKRWIDKENTTIIVDAGKSEDIKARVKQIRAOIETSSDY 360
DB 301 ITEDGIGKLESVTLDMGQAEKVVDKHTTIIGKGDPEQIKARIEQIKRQIQTETSDY 360

QY 361 DREKQERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420
DB 361 DREKQERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420

QY 421 RSKVLDDIKPADDDELAGNIIRRSLEELPQIAANAGYEGSVVEKV-----REPDKGF 476
DB 421 RASEALEDLKGNHDQDGLIIRKAVRPLQIAANAGYEGSVVEKVIELGKEKGVSW 480

QY 477 GNAASGEYEDLTKAGVIDPKKVTIRALQNAASVASLLLTTECAIAEKPEPKK-----D 530
DB 481 GNAATGEYVDMVEAGIIDFTKVVRTAENASVAGTMTAEALIAADLPEKKKIDITPTD 540

QY 531 MP 532
DB 541 MP 542

RESULT 15
US-10-369-493-17132
; Sequence 17132, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17132
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17132

Query Match 65.7%; Score 1780.5; DB 12; Length 544;
Best Local Similarity 64.8%; Pred. No. 3.8e-132;
Matches 355; Conservative 89; Mismatches 95; Indels 9; Gaps 4;

QY 3 SKEILFDKAREKLSRGVDKLANAVKVTGLGKGRNVLEKFGSPVITKDGVSVAKEI 62

DB 2 AKDIKFSDEARRSMRLRGVDKLANAVKVTGLGKGRNVLEKFGSPVITKDGVSVAKEI 61
QY 63 EDKFNMGAMQVEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 122
DB 62 EDAPENMGAKLVAAEASVASKTNDIAGDGTATVLAQAIYREGVKLVAAAGRNPMIAKRGIDK 121

QY 123 AVAVVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAABAMAKVGKGVITVEEAKG 182
DB 122 ATQVAVEELSKTSKIEGKDSIAQVAAS-SADDEVGKIIAEAMERVNGDGVITVEESKG 180

QY 183 LETTLDDVVEGKMPDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKDMPLILEQV 242
DB 181 FSTELEVVVEGKMPDRGYLSPYFVTNPEKMCVCELDNPIYLCNEKKITSMKDMPLILEQV 240

QY 243 VNRPELLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLGGEAIF 302
DB 241 QGKPELLIADVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLGGEAIF 300

QY 303 EDRGIKLENVSLSSLTGAKRWIDKENTTIIVDAGKSEDIKARVKQIRAOIETSSDYDR 362
DB 301 EDGLDLKKSANITQLGRASKVVVTENTTIIVDAGKSEDIKARVKQIRAOIETSSDYDR 360

QY 363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFVS 422
DB 361 EKLQERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFVS 420

QY 423 IKVLDDIKPADDDELAGNIIRRSLEELPQIAANAGYEGSVVEKVVEKPKDGFNAAS 482
DB 421 IKAVSSI-GAGDEATGVNIVLRALEEPVQIAHAGLEGSVIVERLKEEAGFGNAAT 479

QY 483 GEYEDLIKAGVIDPKKVTIRALQNAASVASLLLTTECAIAEKPEPKKDMPEPGG---MG 539
DB 480 GEWVNVVVEAGIIVDPKVTIRALQNAASVASAMFLITTEAVIADKPEENEG---GGCMPDMG 535

QY 540 GNGMGDMG 547
DB 536 GNGMGDMG 543

Search completed: January 28, 2004, 13:22:58
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:11:55 ; Search time 21 seconds
(without alignments)
2509.545 Million cell updates/sec

Title: US-09-077-574A-2

Perfect score: 2708

Sequence: 1 MASKEILFDKAREKLSRGV.....KDMPPGGMGGMGGMDGMY 548

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76.*

1: pirl.*

2: pirl2.*

3: pirl3.*

4: pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	68.9	551	2 JC2562	chaperonin groELx
2	1864	68.8	545	2 C95311	groEL2 chaperonin
3	1863	68.8	545	2 JN0509	heat shock protein
4	1862.5	68.8	547	2 F82783	60kDa chaperonin X
5	1852.5	68.4	546	2 B47073	chaperonin GroEL -
6	1845	68.1	550	2 A41468	60K heat shock pro
7	1844.5	68.1	546	2 S65596	heat shock protein
8	1843.5	68.1	547	2 I40331	Cpn60 protein (Gro
9	1843.5	68.1	552	2 S39765	chaperonin 60 - Co
10	1841	68.0	542	2 F95967	probable heat shoc
11	1840.5	68.0	544	2 AD2660	60 KDa chaperonin
12	1840.5	68.0	544	2 B97442	60K chaperonin (pr
13	1828.5	67.5	544	2 S23918	groEL protein - Ag
14	1825.5	67.4	547	2 B83098	GroEL protein PA43
15	1823	67.3	546	2 S22347	groEL - Brucella a
16	1823	67.3	550	2 S35309	heat shock protein
17	1822	67.3	546	2 I40342	heat shock protein
18	1817.5	67.1	544	2 B36917	heat shock protein
19	1816.5	67.1	541	2 JN0512	heat shock protein
20	1814.5	67.0	544	2 H81964	chaperonin 60kD su
21	1814.5	67.0	549	2 JC2564	heat shock protein
22	1814	67.0	546	2 AG3640	60K chaperonin gro
23	1812.5	66.9	545	2 S47530	chaperonin groEL -
24	1810.5	66.9	544	2 S37039	groEL protein - Ba
25	1808.5	66.8	544	2 C81021	chaperonin, 60 kDa
26	1805.5	66.7	547	2 B43606	heat shock protein
27	1802.5	66.6	544	2 S61300	heat shock protein
28	1799	66.4	546	2 B54539	heat shock protein
29	1794.5	66.3	544	2 S61302	heat shock protein

30 1794 66.2 546 2 S35311 heat shock cognate
31 1791.5 66.2 547 2 B87334 chaperonin, 60 kDa
32 1791 66.1 541 2 S61301 heat shock protein
33 1789.5 66.1 541 2 JN0511 heat shock protein
34 1788.5 66.0 544 2 S61303 heat shock protein
35 1786 66.0 545 2 C70489 groEL - Aquifex ae
36 1780.5 65.7 544 2 JC06063 chaperonin groEL -
37 1780.5 65.7 544 2 B83720 class I heat-shock
38 1774.5 65.5 545 2 S51563 heat shock protein
39 1773 65.5 544 1 B43827 chaperonin groEL -
40 1765 65.2 546 2 S34938 heat shock protein
41 1761.5 65.0 544 2 B82048 chaperonin, 60 kD
42 1761 65.0 548 2 AG0043 60 kDa chaperonin
43 1759 65.0 548 2 C64076 chaperonin groEL -
44 1755.5 64.8 539 2 B49855 heat shock protein
45 1755 64.8 545 2 G81328 60 kD chaperonin (

ALIGNMENTS

RESULT 1

JC2562

chaperonin groELx protein - Amoeba proteus

C:Species: Amoeba proteus

C:Date: 17-May-1995 #sequence_revision 14-Jul-1995 #text_change 26-Aug-1999

C:Accession: JC2562

R:Alm, T.I.; Lim, S.T.; Lee, H.K.; Jeon, K.W.

Gene 148[128], 43-49, 1994

A:Title: A novel strong promoter of the groEx operon of symbiotic bacteria in Amoeba prot

A:Reference number: JC2561

A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2562

A:Molecule type: DNA

A:Residues: 1-551 <AHN>

A:Cross-references: GB:M86549; NID:G155400; PIDN:AAC09381.1; PID:G155402

C:Comment: This protein is involved in the assembly of oligomeric protein complexes, and

C:Genetics:

A:Gene: groELx

C:Superfamily: chaperonin groEL

C:Keywords: molecular chaperone

Query Match 68.9%; Score 1866; DB 2; Length 551;

Best Local Similarity 67.9%; Pred. No. 2.2e-83;

Matches 373; Conservative 74; Mismatches 96; Indels 6; Gaps 2;

Qy 3 SKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSFSGSPVITKDGVSVAKEIEL 62

Db 2 AKELRFGDDARQQMLAGVNLADRVKATMGPSGRNVVLSRSGAPTPTVKDGVSVAKEIEF 61

Qy 63 EDKFNNGAQMVEVAPKTSIAGDGTATVLAQAIYREGVKLVAAGRNPMKRGIDK 122

Db 62 ENRFKNNGAQMVEVAPKTSIAGDGTATVLAQAIYREGVKLVAAGRNPMKRGIDK 121

Qy 123 AVAVTKELSDITKPTDQKEIAQVGTISANSDDTGNIIIAEMAKVKGCVITVEAKG 182

Db 122 AVTAITKELQMSKPKCKGKAIQVGTISANSDDTGNIIIAEMAKVKGCVITVEAKG 181

Qy 183 LETTLDVVEGMKFDKRGYLSFYFVTNPKMVCLELDNPVILCNKKITSMKDWLPLEQVAK 242

Db 182 LENELSVVEGMKFDKRGYLSFYFVTNPKMVCLELDNPVILCNKKITSMKDWLPLEQVAK 241

Qy 243 VNRPLIIAEDVEGEMALATLVNKLKRGALQVAVKAPGFGERRKAMLEDITAILTGGBAIF 302

Db 242 SGRPLIIAEDVEGEMALATLVNKLKRGALQVAVKAPGFGERRKAMLEDITAILTGGBAIF 301

Qy 303 EDGKIKLENVLSLGTAKRVVDKENTTIVDCAGKSEDIKARVKQIRAQIBETSSDYDR 362

Db 302 EETGTSLETASLESGLTAKRVVDKENTTIVDCAGKSEDIKARVKQIRAQIBETSSDYDR 361

Qy 363 EKLQERLAKLVGGVAHVHGAATETEMKKKRDVEDALNATRAAVEGIVPGGTAFTVRS 422

Db 362 EKLQERLAKLVGGVAHVHGAATETEMKKKRDVEDALNATRAAVEGIVPGGTAFTVRS 421

```
QY 423 IKVLDDIKPADDDELALAGNIIRSRLEELRQIAANAGYEGSVVEKREPKDGFQFNAAS 482
Db 422 QKVLDDGLKGNADQDMGINILRRATESPLRQIVANAGYESSVIVNKVAEHKDNFQFNAAT 481
QY 483 GYEDLIKAGVTDPKKVTTRIALQNAASVASILLTTTECAIAEKPEPKKOMPMPG----GSM 538
Db 482 GOYGDVEMGLIDPKVTRTALQNAASVRSMLTTECVADL--PKDDEGAGAGDMGSM 539
QY 539 GGMGMDGM 547
Db 540 GGMGMDGM 548

RESULT 2
C95311
groEL2 chaperonin [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSYMA
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: C95311
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: C95311
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-545 <R>
A:Cross-references: GB:AE006469; PIDN:AAK65053.1; PID:g14523485; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSYMA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, B.; Komp, C.; Lelaure,
hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: groEL2
A:Genome: plasmid
C:Superfamily: chaperonin groEL

Query Match 68.8%; Score 1864; DB 2; Length 545;
Best Local Similarity 70.5%; Pred. No. 2.7e-83;
Matches 387; Conservative 59; Mismatches 97; Indels 6; Gaps 4;

QY 1 MASKEILFDAKAREKLSRGVDKLANAVKVTGLGPKGNVVIKESFGSPVITKGVSAKEI 60
Db 1 MAAKEVKFGSAREKMLRGVDILADAVKVTGLGPKGNVVIDKSGFAPRITKGVSAKEI 60
QY 61 ELEDKFNMGAAQMVKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAAGNPMIAIKRGI 120
Db 61 ELEDKFNMGAAQMVREVASKTNDIADGTTTATVLAQAIYREGAKAQAAGNPMIAIKRGI 120
QY 121 DKAVAVATKELSDITKPTROQKEIAQVGTTSANSDDTTIGNIIAAMAKVKGKGVITVEEA 180
Db 121 DLAAEVVKDILLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEA 180
QY 181 KGLTTLDDVVEGKMPDRGYLSPYFVTNPKWVCNLPYLLCNEKKTSMKMLPILEQV 240
Db 181 KTAETELVVEGQFDRGYLSPYFVTNPKWVADLEDAFILLHEKKSINLQAMLPVLEAV 240
QY 241 AKVNRPLIIIAEDVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGGEA 300
Db 241 VQTKPELLIIAEDVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGTV 300
QY 301 IFEDRGKLENVLSISLGTAKRVVIDKENTTIIVDGAKSSEDIKARVKQIAQIETSSDY 360
Db 301 ISEDLGKLESVTLDMGRKAKVITKENTTIIVDGAKQKSDIEGRVAQIAQIETSSDY 360
QY 361 DREKLERLAKVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLERLAKVGGVAVIRVGGATEVEVEKKEKKORIDALNATRAAEEGIVPGGGVALL 420
QY 421 R-SIKVLDLDDIKPADDDELALAGNIIRSRLEELRQIAANAGYEGSVVEKREPK-DGFGF 478
Db 421 RSSVKI--TVKGENDDQDAGVIVRRALQSQPARQIVENAGDEASIVVGKILEKNTDDFGY 478
```

```
QY 361 DREKLERLAKVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLERLAKVGGVAVIRVGGATEVEVEKKEKKORIDALNATRAAEEGIVPGGGVALL 420
QY 421 R-SIKVLDLDDIKPADDDELALAGNIIRSRLEELRQIAANAGYEGSVVEKREPK-DGFGF 478
Db 421 RSSVKI--TVKGENDDQDAGVIVRRALQSQPARQIVENAGDEASIVVGKILEKNTDDFGY 478
QY 479 NAASEYEDLIKAGVIDPKKVTTRIALQNAASVASILLTTTECAIAEKPEPKKOMPMPGGM 538
Db 479 NAAQTEGYDMTANGIIDPKVTRTALQNAASVASILLITTEMAEL--PKKDAPAMPGM 536
QY 539 GGMGMDGM 547
Db 537 GGMGMDGM 545

RESULT 3
JN0509
heat shock protein groEL (clone Rhz A) - Rhizobium meliloti
N:Alternate names: chaperonin groEL protein
C:Species: Rhizobium meliloti
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: JN0509
R:Rusanganwa, E.; Gupta, R.S.
Gene 126, 67-75, 1993
A:Title: Cloning and characterization of multiple groEL chaperonin-encoding genes in Rh
A:Reference number: JN0509; MUID:93231539; PMID:8097179
A:Accession: JN0509
A:Molecule type: DNA
A:Residues: 1-545 <R>
A:Cross-references: GB:M94192; NID:g152233; PIDN:AAA26285.1; PID:g152235
C:Comment: This protein plays a role in protein folding and in the extracellular transpo
C:Genetics:
A:Gene: groEL
A:Superfamily: chaperonin groEL
C:Keywords: heat shock; molecular chaperone; stress-induced protein

Query Match 68.8%; Score 1863; DB 2; Length 545;
Best Local Similarity 70.5%; Pred. No. 3e-83;
Matches 387; Conservative 58; Mismatches 98; Indels 6; Gaps 4;

QY 1 MASKEILFDAKAREKLSRGVDKLANAVKVTGLGPKGNVVIKESFGSPVITKGVSAKEI 60
Db 1 MAAKEVKFGSAREKMLRGVDILADAVKVTGLGPKGNVVIDKSGFAPRITKGVSAKEI 60
QY 61 ELEDKFNMGAAQMVKEVAPKTSIADGTTTATVLAQAIYREGVKLVAAAGNPMIAIKRGI 120
Db 61 ELEDKFNMGAAQMVREVASKTNDIADGTTTATVLAQAIYREGAKAQAAGNPMIAIKRGI 120
QY 121 DKAVAVATKELSDITKPTROQKEIAQVGTTSANSDDTTIGNIIAAMAKVKGKGVITVEEA 180
Db 121 DLAAEVVKDILLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEA 180
QY 181 KGLTTLDDVVEGKMPDRGYLSPYFVTNPKWVCNLPYLLCNEKKTSMKMLPILEQV 240
Db 181 KTAETELVVEGQFDRGYLSPYFVTNPKWVADLEDAFILLHEKKSINLQAMLPVLEAV 240
QY 241 AKVNRPLIIIAEDVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGGEA 300
Db 241 VQTKPELLIIAEDVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDAIILTGTV 300
QY 301 IFEDRGKLENVLSISLGTAKRVVIDKENTTIIVDGAKSSEDIKARVKQIAQIETSSDY 360
Db 301 ISEDLGKLESVTLDMGRKAKVITKENTTIIVDGAKQKSDIEGRVAQIAQIETSSDY 360
QY 361 DREKLERLAKVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLERLAKVGGVAVIRVGGATEVEVEKKEKKORIDALNATRAAEEGIVPGGGVALL 420
QY 421 R-SIKVLDLDDIKPADDDELALAGNIIRSRLEELRQIAANAGYEGSVVEKREPK-DGFGF 478
Db 421 RSSVKI--TVKGENDDQDAGVIVRRALQSQPARQIVENAGDEASIVVGKILEKNTDDFGY 478
```

[illegible]

RESULT 5

B47073
Chaperonin GroEL - Chromatium vinosum
C:Species: Chromatium vinosum
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C:Accession: B47073
R:Rerrevira, R.G.; Soncini, F.C.; Viale, A.M.
J. Bacteriol. 175, 1514-1523, 1993
A:Title: Cloning, characterization, and functional expression in Escherichia coli
A:Reference number: A47073; PMID:8444812
A:Accession: B47073
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-546 <P>
A:Cross-references: GB:M99443; NID:gl45007; PIDN:AAA23319.1; PID:gl45009
A:Note: sequence extracted from NCBI backbone (NCBIN:126968, NCBIP:126970)
C:Superfamily: Chaperonin GroEL
C:Keywords: molecular chaperone

Query Match 68.4%; Score 1852.5; DB 2; Length 546;
Best Local Similarity 68.2%; Pred. No. 9.7e-83;
Matches 374; Conservative 70; Mismatches 101; Indels 3; Gaps 2;

1 MASKEILFDKAREKLRSQVKLANAVVTLGPKGRNVVIEKSFSGSVITKOGVSVAKEI 60
:::
1 MSADVKGFGDARVRMMEGVNILANAVVTLGPKGRNVVLEKSFAGATVTWKOGVSVAKEI 60

61 ELEDKFENWGAQMVKEVAPKTSIDAGDGTTTATVLAQAITYRGVKI VAAGRNPMAIKRGI 120

61 ELKDFENWGAQMVKEVAKTSIDAGDGTTTATVLAQWVYRGKLKVAAGMPPMDIKRGM 120

121 DKAVVATKELSDITKPTRDQKEIAQVTISANSDDTTCIIIAEAMAKYKGCVITVEEA 180
||||| :|| :|| ||||| :|| ||||| :|| ||||| :||
121 DKAVEATEELKKLSKPCPRPWAQAQVTISANSDDSGTCTIAEAMEKYKGCVITVEDG 180

181 KGLETTLDVVEGMKFDRCGLSPFYFTNPEKMWCELDNPYILCNEKKITSMKMDLPLEQV 240
 - : |||||:|||||:|||||: : |||||:|||||:|||||:|||||:|||||:|||||:
 181 TSLQNELDVVEGMQFDRCGLSPFYFNQQSQSAELDAPYILLYDKKISNIRULLPVEGV 240

241 AKVNRPLIIAEDVEGEALATLVNKIRGALQVVAVKAPGGERRRKMLEDIAITGTGSEA 300
|| : ||| ||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 AKAGKPILLI AEDVEGEALATLVNTRTGRIVKVCAVKAPGGDRRRKLQDIAITGATGV 300

```

301 IFEDRGIKLENNSSIGTAKRVVIDKENTIVDGAGKSEIDIKARVQIQRAQIETSSY 360
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
301 ISEVGLSLKEATUDTGTAQRVQGKDFTIIDGSGSEIIDKACEQIRAQVETSSY 360

```

[illegible]

```

421 RSTKVLDDIKPADDELAGLNIIRRSLEELPQIAANAGYEGSVIVEKVPKDGFGNA 480
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 RAIAAVKDLKGANHDQVGTIAIRRAEELPQIVANAGREPSVTIHKVACTGTGNGNA 480

```

481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLTTTECAIAEKPEPKDMP-MFGGGMG 539

Qy	479	NAASGEYDLIKAGVIDPKKVTRIALQNAASVASLILITTECAIAEPEPKKMPEGGM	538
Db	479	NAQTGEYGDMIAMGIIDPKKVVRTALQDAASVASLILITTEAMIAEL--PKKDAPAMEGGM	536
Qy	539	GGMGGMGDM	547
Db	537	GGMGGMMDM	545

RESULT 4

F82783
 60kDa chaperonin XP0615 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: F82783
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: F82783
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-547 <SIM>
 A:Cross-references: GB:AE003907; GB:AE003849; NID:g9105484; PIDN:AAF83425.1; GSPDB:GN001
 A:Experimental source: strain 9a5c

A.; Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Arantes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Docena, C.; El-Dorzy, H.; Faicincani, A.P.; Ferreira, A.J.S.

Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Maraca, E.C.; Minaki, C.V.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
Tsuchako, M.H.; VALLADA, H.; Van Sluys, M.A.; Viorioveski, Almeida, C.

[illegible]

Query Match 68.8%; Score 1862.5; DB 2; Length 547;
Best Local Similarity 66.8%;
Superfamily: chaperonin groEL

[illegible]

bb
1 MAAKEIIFSEKARSRMHGVNLLANAVKATIGPKGRHVLDKSPITIKDGVSAKEI 60
61 ELEDKFNMGAAQWKEVAPKTSDIAGDQTTATVLAQAIYREGVKLYAAGENPMAIKRG 120

61. ELADKFENMGAQMLKEYASTNDHAGDTTATVLAQALIREGCKVAAGNNPDLRGI 120

121 DKAAVAVTKELSDITKPTRDOKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEEA 180

121 DKAVIAAVTELKKISKSTPSDDKATAQVATISANSDESIGNIIAEAMKKVKGEGVITTEEG 180
181 KGLETTLLDVVEGKKFDRGYLSPYFVTVNPERKVMCELDNPPYILCNEKKITSMKMDLPILLEQV 240

181 TTLENELDVVEGMOQFDRGYSPYFINNQSQIVELDNPFYILLHDKKISSVRDLLTVLDAV 240

241 AKVNRPLIIITAEDEVEGALATLVNKLRGALQVAVKAPGERRKKAWLEJAITLTGEA 300

241 AKESKPLLIIVAAEEVEGALATLVNNIRGIKVCAPKPGFDRRKAMLEDMAVLTGTV 300

301 IFEDRGIKLENVLSLSIGTAKRVVIDKENTTIVDGAKSSEDIKARVKQIRAIQIEETSSDY 360

301 ISEEVGLSEKATTSHLGKAKKVRVSKENTTIIDGIGDNDAINGRVKIQTQIEETTSY 360

Db 481 ANGEYGVNEMGILDPKTRVTSALONSCSVAGLMTTTEAMIAD-EPKDDAPAMPGGGMG 538
 QY 540 GMGGMGDM 547
 Db 539 DMGGMGM 546

RESULT 6
 A41468
 60K heat shock protein htpB - Legionella pneumophila
 C:Species: Legionella pneumophila
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
 C:Accession: A41468; A41472
 R:Hofman, P.S.; Houston, L.; Butler, C.A.
 Infect. Immun. 58, 3380-3387, 1990
 A:Title: Legionella pneumophila htpAB heat shock operon: nucleotide sequence and expression
 A:Reference number: A41468; MUID:90382960; PMID:220580
 A:Accession: A41468
 A:Molecule type: DNA
 A:Residues: 1-550 <HOF>
 A:Cross-references: GB:M31918
 R:Samson, J.S.; O'Connor, S.P.; Holloway, B.P.; Plikaytis, B.B.; Carlone, G.M.; Mayer, Infect. Immun. 58, 3154-3157, 1990
 A:Title: Nucleotide sequence of htpB, the Legionella pneumophila gene encoding the 58-kDa
 A:Reference number: A41472; MUID:90354095; PMID:2117582
 A:Accession: A41472
 A:Molecule type: DNA
 A:Residues: 3-188, 'S', 190-195, 'LIAVH', 202-347, 'T', 349-550 <SAM>
 A:Cross-references: GB:M91673; GB:M35149; NID:g149689; PIDN:AAA25298.1; PID:g149690
 C:Genetics:
 A:Gene: htpB
 C:Superfamily: chaperonin groEL

Query Match 68.1%; Score 1845; DB 2; Length 550;
 Best Local Similarity 67.3%; Pred. No. 2.3e-82;
 Matches 367; Conservative 80; Mismatches 96; Indels 2; Gaps 1;
 QY 3 SKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSPGSPVITKDGVSVAKEI 62
 Db 4 AKELRGDDARLQMLAGVNALADAVQWTGPRGRNVVIEKSPGSPVITKDGVSVAKEI 63
 QY 63 EDKFENMGQMKVEAPKSDIAGDGTATVLAQAIYREGVKLVAAGRNPMKRGIDK 122
 Db 64 EHRFMNGAQMKVEAVASKSDTAGDGTATVLAQAIYREGVKLVAAGRNPMKRGIDK 123
 QY 123 AVAVATKELSDITKPTRDQKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEAKG 182
 Db 124 AVLAVTKLQAMSKPKCKDSKAIQVGTISANSDEAIGALIAEAMKVGKGVITVEDNG 183
 QY 183 LETTLDVVGKMPDRGYLSPYFVTNPEKQVCELDNPIYLCNEKKITSMKMDLPILQVAK 242
 Db 184 LENELVVGKMPDRGYLSPYFVTNPEKQVCELDNPIYLCNEKKITSMKMDLPILQVAK 243
 QY 243 VNPPLIIAEDVVEGEALATLVNKLRGALOVVAVKAPGFGERRKAMLEDIAILLTGGAIF 302
 Db 244 SGRPLIIAEDVVEGEALATLVNKLRGALOVVAVKAPGFGERRKAMLEDIAILLTGGAIF 303
 QY 303 EDRGKLENVSLSLGTAKRVVIDKENTTIIVDGAGKSDIIRKVKQIRAOIEETSSDYR 362
 Db 304 EELGKLENTLSDGSAKRVVTKENTTIIVDGAGKSDIIRKVKQIRAOIEETSSDYR 363
 QY 363 EKLQERLAKLVGGVAVIHVGAATEMKEKDRVEDALNATRAAVERGIVPGGTAFAVS 422
 Db 364 EKLQERLAKLVGGVAVIHVGAATEMKEKDRVEDALNATRAAVERGIVPGGTAFAVS 423
 QY 423 IKVLDDIKPADDDDELAGNIIRRSLEELPQIAANAGVEGSIIVVEKREPDKGFGNAAS 482
 Db 424 QKALDSLKGNDDQNMGINIIRRSLEELPQIAANAGVEGSIIVVEKREPDKGFGNAAS 483
 QY 483 GEYEDLIKAGVIDPKKVTIRALQNAASVASLLLTTECAIAEKPBPMPGGMGMG 542
 Db 484 GEYGVNEMGILDPKTRVTSALONSCSVAGLMTTTEAMIAD-EPKDDAPAMPGGGMG 541

QY 543 GMGDM 547
 Db 542 GMGGM 546

RESULT 7
 S65596
 heat shock protein 60 - Rhizobium leguminosarum
 N:Alternate names: chaperonin 60
 C:Species: Rhizobium leguminosarum
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
 C:Accession: S65596
 R:Wallington, E.J.; Lund, P.A.
 Microbiology 140, 113-122, 1994
 A:Title: Rhizobium leguminosarum contains multiple chaperonin (cpn60) genes.
 A:Reference number: S65596; MUID:94214663; PMID:7909257
 A:Accession: S65596
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-546 <WAL>
 A:Cross-references: EMBL:L20775; NID:g387873; PIDN:AAA26246.1; PID:g387874
 C:Genetics:
 A:Gene: cpn60
 C:Superfamily: chaperonin groEL
 C:Keywords: heat shock; molecular chaperone

Query Match 68.1%; Score 1844.5; DB 2; Length 546;
 Best Local Similarity 69.8%; Pred. No. 2.4e-82;
 Matches 384; Conservative 61; Mismatches 98; Indels 7; Gaps 5;
 QY 1 MASKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 Db 1 MASKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 QY 61 ELEDKFPENMGQMKVEAPKSDIAGDGTATVLAQAIYREGVKLVAAGRNPMKRGIDK 120
 Db 61 ELEDKFPENMGQMKVEAPKSDIAGDGTATVLAQAIYREGVKLVAAGRNPMKRGIDK 120
 QY 121 DKAVAVATKELSDITKPTRDQKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEEA 180
 Db 121 DLAVAVATKELSDITKPTRDQKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEEA 180
 QY 181 KGLETTLDVVGKMPDRGYLSPYFVTNPEKQVCELDNPIYLCNEKKITSMKMDLPILQV 240
 Db 181 KTAFTLELVVGKMPDRGYLSPYFVTNPEKQVCELDNPIYLCNEKKITSMKMDLPILQV 240
 QY 241 AKVNRPLIIAEDVVEGEALATLVNKLRGALOVVAVKAPGFGERRKAMLEDIAILLTGTV 300
 Db 241 VQTGKPLIIAEDVVEGEALATLVNKLRGALOVVAVKAPGFGERRKAMLEDIAILLTGTV 300
 QY 301 IFEDRGKLENVSLSLGTAKRVVIDKENTTIIVDGAGKSDIIRKVKQIRAOIEETSSDY 360
 Db 301 ISEDLGKLENVSLSLGTAKRVVIDKENTTIIVDGAGKSDIIRKVKQIRAOIEETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGAATEMKEKDRVEDALNATRAAVERGIVPGGTAFAV 420
 Db 361 DREKLOERLAKLVGGVAVIHVGAATEMKEKDRVEDALNATRAAVERGIVPGGTAFAV 420
 QY 421 RSKVLDDIKPADDDDELAGNIIRRSLEELPQIAANAGVEGSIIVVEKREPDKGFGFN 479
 Db 421 RSKTKI-TVKGANDDQAGINIVRRALQSLVRQIAENAGDEASIVGKGLDKNEDNFGYN 479
 QY 480 AASEYEDLIKAGVIDPKKVTIRALQNAASVASLLLTTECAIAEKPBPMPGGMGMG 537
 Db 480 AQTSEYEDLIKAGVIDPKKVTIRALQNAASVASLLLTTECAIAEKPBPMPGGMGMG 536
 QY 538 MGGGMGDM 547
 Db 537 MGGGMGDM 546

RESULT 8

I40331
Cpn60 protein (GroEL) - Bordetella pertussis
C:Species: Bordetella pertussis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 26-Aug-1999
C:Accession: I40331
R;Fernandez, R.C.; Weiss, A.A.
Gene 158, 151-152, 1995
A:Title: Cloning and sequencing of the Bordetella pertussis cpn60 (groESL) homolog
A:Reference number: I40330; MUID:95309719; PMID:7789805
A:Accession: I40331
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-547 <RES>
A:Cross-references: EMBL:U12277; NID:g968918; PIDN:AAA74567.1; PID:g968920
C:Genetics:
A:Gene: cpn60
C:Superfamily: chaperonin groEL

Query Match 68.1%; Score 1843.5; DB 2; Length 547;
Best Local Similarity 67.4%; Pred. No. 2.7e-82;
Matches 370; Conservative 74; Mismatches 98; Indels 7; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGPKGRNVVIEKSGFSPVITKDGVSVAKEI 60
DB 1 MAAKQVLFDAEARVIRVGVNVLNANAVKVTGLGPKGRNVVLEKSGFAPTVTKDGVSVAKEI 60
QY 61 ELEDKFNNGAQMVKVAPKTSIDAGDGTATVLAQAIVREGVKLVAAAGRNPMIAIKRGI 120
DB 61 ELKDFENIGAQVVKDASVTSNAGDGTATVLAQAIVREGVKLVAAAGRNPMIAIKRGI 120
QY 121 DRKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEEA 180
DB 121 DRKAVAAVELKKSIPVTSKEIAQVGSISANSASIGIADAMDKVKGKGVITVEEDG 180
QY 181 KGLETTLDVVEGGMKFDGKYLSPYFVTNPEKMWCELDNPNYILCNEKKITSMKMLPILVLEQV 240
DB 181 KGLETTLDVVEGGMKFDGKYLSPYFINSPEKQVAALDDPYVLIYDKKVSNIKRLDLPVLEQV 240
QY 241 AKVNRPLLIIEADEVGEALATLVNKLQALGVAVKAPGFGERRKAMLEIDAILTGGEA 300
DB 241 AKSSRPLLIIEADEVGEALATLVNKNIRGILKTTAVKAPGFGERRKAMLEIDAILTGGEV 300
QY 301 IFEDRGKILENYSLSIGTAKRVVIDKENTTIIDGAGKSEDIKARVKQIRAEIEETSSDY 360
DB 301 ISEETGMSLEKATLQDLQAKRIEVAKENTTIIDGAGDKSIEARVKQIRAEIEETSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATEMKEKDRVEDALNATRAAIVEGIVPGGGTAFV 420
DB 361 DREKLOERLAKLVGGVAVIRVGAATEVENKEKARVEDALHATRAAIVEGIVPGGGVALL 420
QY 421 RSKVLDDIKPADDDELAGNIIRRSLEPLRQIAANAGYEGSVIVVEKREPKDGFQFNA 480
DB 421 RAKQAITGLKGTADQNAQIKLIRAVEPLRTIVTNGAGDEASVVVNTVNLKGNTRYNA 480
QY 481 ASGEYEDLIKAGVIDPKKVTIRALQNAASVASLLITTECAIAE---KPEPKKDMPPGG 536
DB 481 ATGEYGLVVEQGVLDPTKVTIRALQNAASVASLLITTECAIAE---KPEPKKDMPPGG 537
QY 537 GMGGMGMD 545
DB 538 GMGGMGMD 546

RESULT 9
S39765
chaperonin 60 - Coxiella burnetii
C:Species: Coxiella burnetii
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 28-May-1999
C:Accession: S39765
R;Vodkin, M.H.; Williams, J.C.
J. Bacteriol. 170, 1227-1234, 1988
A:Title: A heat shock operon in Coxiella burnetii produces a major antigen homologous to
A:Reference number: S39764; MUID:88139182; PMID:3343219

A:Accession: S39765
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <VOD>
A:Cross-references: GB:M20482; NID:g144996; PIDN:AAA23309.1; PID:g144998
C:Superfamily: chaperonin groEL

Query Match 68.1%; Score 1843.5; DB 2; Length 552;
Best Local Similarity 66.5%; Pred. No. 2.7e-82;
Matches 365; Conservative 82; Mismatches 99; Indels 3; Gaps 3;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGPKGRNVVIEKSGFSPVITKDGVSVAKEI 60
DB 1 MAAKVLKFSHEVLHMSRGVEVLNANAVKVTGLGPKGRNVLDKSGFAPTVTKDGVSVAKEI 60
QY 61 ELEDKFNNGAQMVKVAPKTSIDAGDGTATVLAQAIVREGVKLVAAAGRNPMIAIKRGI 120
DB 61 ELEDKFNNGAQMVKVAPKTSIDAGDGTATVLAQAIVREGVKLVAAAGRNPMIAIKRGI 120
QY 121 DRKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEEA 180
DB 121 DRKAVAAVELKKSIPVTSKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEEDG 180
QY 181 KGLETTLDVVEGGMKFDGKYLSPYFVTNPEKMWCELDNPNYILCNEKKITSMKMLPILVLEQV 240
DB 181 KGLETTLDVVEGGMKFDGKYLSPYFINSPEKQVAALDDPYVLIYDKKVSNIKRLDLPVLEQV 240
QY 241 AKVNRPLLIIEADEVGEALATLVNKLQALGVAVKAPGFGERRKAMLEIDAILTGGEA 300
DB 241 AKSSRPLLIIEADEVGEALATLVNKNIRGILKTTAVKAPGFGERRKAMLEIDAILTGGEV 300
QY 301 IFEDRGKILENYSLSIGTAKRVVIDKENTTIIDGAGKSEDIKARVKQIRAEIEETSSDY 360
DB 301 ISEEVGLSLEAASLDLGSARKRVVVTDDTTIIDGSDAGDIKNRVEQIRKEIENSSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATEMKEKDRVEDALNATRAAIVEGIVPGGGTAFV 420
DB 361 DREKLOERLAKLVGGVAVIRVGAATEVENKEKARVEDALHATRAAIVEGIVPGGGVALL 420
QY 421 RSKVLDDIKPADDDELAGNIIRRSLEPLRQIAANAGYEGSVIVVEKREPKD-GFGFN 479
DB 421 RVLKSLDSSEVENEDQRVGEIARRAAYPLSGIVKNTGVQAAVADVADKVLNHNKGVN 480
QY 480 ASGEYEDLIKAGVIDPKKVTIRALQNAASVASLLITTECAIAEKEPKKDMPPGG-M 538
DB 481 AATGEYGLVVEQGVLDPTKVTIRALQNAASIASIAGMITTECMVTEAPK-KKEESMPGGMD 539
QY 539 GMGGMGMD 547
DB 540 GMGGMGMD 548

RESULT 10
F95967
probable heat shock protein groEL [imported] - Sinorhizobium meliloti (strain 1021) magaf
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: F95967
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernand
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endos
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: F95967
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-542 <MUR>
A:Cross-references: GB:AL591985; PIDN:CAC49406.1; PID:g15140892; GSPDB:GNO0167
A:Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, P.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenbol, M.; Vortholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: groEL5; SMD21566
A:Genome: plasmid
C:Superfamily: chaperonin groEL

Query Match 68.0%; Score 1841; DB 2; Length 542;
Best Local Similarity 68.3%; Pred. No. 3.5e-82;
Matches 373; Conservative 71; Mismatches 96; Indels 6; Gaps 2;
QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTI L GPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
DB 1 MAAREVKFQDAREKMLRGVDVLANAVKVTI L GPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
QY 61 ELEDKFENMGQAMKREVAAPKTSIAGDGT T TATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
DB 61 ELEDKFENMGQAMKREVAAPKTSIAGDGT T TATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
QY 121 DKAVVAVTKELSDITKPTRDQKEIAQVGTI SANS DTTTGNIIAEAMAKVKGKGVITVEEA 180
DB 121 DKAVVAVTKELSDITKPTRDQKEIAQVGTI SANS DTTTGNIIAEAMAKVKGKGVITVEEA 180
QY 181 KGLETTLDVVGKMFDRGYLSPFYFVTNPEKMWCELDNPIYLCNEKKITSMKMDLPILBOV 240
DB 181 KTAETELEVVGMDRGYLSFYFVTNPEKMWCELDNPIYLCNEKKITSMKMDLPILBOV 240
QY 241 AKVNRPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEIDIAITGGTGA 300
DB 241 VQSGKPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEIDIAITGGTGA 300
QY 301 IFEDRGIKLENVSLSLGTAKRVVIDKENTTIVD GAGKSEDIKARVKQIRAIQIETSSDY 360
DB 301 VSEDIGIKLESVTLMDLGRKKVSIKENTTIVD GAGKSEDIKARVKQIRAIQIETSSDY 360
QY 361 DREKLOERLAKLVGVAVIHVGAATETEMKEDKRDVEDALNATRAAEEGIVPGGGTAFV 420
DB 361 DREKLOERLAKLVGVAVIHVGAATETEMKEDKRDVEDALNATRAAEEGIVPGGGTAFV 420
QY 421 RSKVLDDIKPADDDDELAGLNIIRSLPELRQIAANAGYEGSI VVEKVRREP-KDGFQFN 479
DB 421 RAVKALDGLKTANNDRQVGLVRAIEAPVQIAENAGYEGSI VVEKVRREP-KDGFQFN 479
QY 480 AASGEYEDLIKAGVIDPKVTRIALQNAASVASL LTTTECAIAEKPEPKKMPMPGGMG 539
DB 480 AACTNEYGDLVAMGVIDPAKVVRTALQDAASVAGL LVTTEAMIAEKPEKKEAAPPALPAG--- 537
QY 540 GMGGMD 545
DB 538 --GMMD 541

RESULT 11
AD2660
60 KDA chaperonin [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD2660
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41698.1; PID:g17739044; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: groEL
A:Map position: circular chromosome
C:Superfamily: chaperonin groEL

Query Match 68.0%; Score 1840.5; DB 2; Length 544;
Best Local Similarity 69.2%; Pred. No. 3.7e-82;
Matches 376; Conservative 69; Mismatches 95; Indels 3; Gaps 3;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTI L GPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
DB 1 MAAREVKFQDAREKMLRGVDVLANAVKVTI L GPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
QY 61 ELEDKFENMGQAMKREVAAPKTSIAGDGT T TATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
DB 61 ELEDKFENMGQAMKREVAAPKTSIAGDGT T TATVLAQAIYREGVKLVAAAGRNPMIAKRG 120
QY 121 DKAVVAVTKELSDITKPTRDQKEIAQVGTI SANS DTTTGNIIAEAMAKVKGKGVITVEEA 180
DB 121 DLAVAEVVKDLQAKAKKINTSEEAQVGTI SANGERQIGLDIAEAMQVNGEGVITVEEA 180
QY 181 KGLETTLDVVGKMFDRGYLSPFYFVTNPEKMWCELDNPIYLCNEKKITSMKMDLPILBOV 240
DB 181 KTAETELEVVGMDRGYLSFYFVTNPEKMWCELDNPIYLCNEKKITSMKMDLPILBOV 240
QY 241 AKVNRPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEIDIAITGGTGA 300
DB 241 VQSGKPLIIAEDVEGEALATLVNKLKGLQVAVKAPGFGERRKAMLEIDIAITGGTGA 300
QY 301 IFEDRGIKLENVSLSLGTAKRVVIDKENTTIVD GAGKSEDIKARVKQIRAIQIETSSDY 360
DB 301 ISEDIGIKLESVTLMDLGRKKVSIKENTTIVD GAGKSEDIKARVKQIRAIQIETSSDY 360
QY 361 DREKLOERLAKLVGVAVIHVGAATETEMKEDKRDVEDALNATRAAEEGIVPGGGTAFV 420
DB 361 DREKLOERLAKLVGVAVIHVGAATETEMKEDKRDVEDALNATRAAEEGIVPGGGTAFV 420
QY 421 RSKVLDDIKPADDDDELAGLNIIRSLPELRQIAANAGYEGSI VVEKVRREP-KDGFQFN 479
DB 421 RSSTKI-TVGVNDQDQAGINIVKALQSLVROIAENAGDEASIVVGKILDKNEDNYGN 479
QY 480 AASGEYEDLIKAGVIDPKVTRIALQNAASVASL LTTTECAIAEKPEPKKMPMPGGMG 538
DB 480 AACTNEYGDLVAMGVIDPAKVVRTALQDAASVAGL LVTTEAMIAEKPEKKEAAPPALPAG--- 539
QY 539 GGM 541
DB 540 GGM 542

RESULT 12
B97442
60K chaperonin (protein cpn60) (groEL protein) [imported] - Agrobacterium tumefaciens (S
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: B97442
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B97442
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86491.1; PID:g15155641; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C 1220
A:Map position: circular chromosome
C:Superfamily: chaperonin groEL
Query Match 68.0%; Score 1840.5; DB 2; Length 544;

181 SLENEISVVEGMDRGYLSPYFVNKPDWAAELDSPLLLVDKISNIREMLPVLAV 240
241 AKVNRPLIIIAEDVEGEALATLVNKLRGALQVAVKAPGGERRKAMLEDAIILTGEA 300
241 AKAGRPILLIAEDVEGEALATLVNKRGIKVAAVKAPGGERRKAMLODIAIILTGTV 300
301 IFEDRGKLENVSSISLGTAKRVVDKENTTTVDGAGKSEDIKARVKQIRAQIBETSSDY 360
301 ISEEVGLSLEGATLEHLGNKRVVINKENTTTIDGAGVQADIEARVLQIRQIBETSSDY 360
361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420
361 DREKLOERLAKLVGGVAVIKVGAATEVEMKEKARVEDALHATRAAEEGVVPGGGVALV 420
421 RSIKVLDDIKPADDDELAGLNIIRSLRQIAANAGYEGSIVVEKRVKDFGFGNA 480
421 RALQALIEGLKGNBQNVGIALRRRAVESPLRQIVANAGDEPSVVVDVKQSGNYGFNA 480
481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEPKDMPPG--GM 538
481 ATGVYGDMEIGILDPAKVTRESALQAAASIGGLMITTEAMVAIVE---DKPAMGMPDM 537
539 GGMGMDGM 547
538 GGMGMDGM 546

RESULT 15
S22347
groEL - Brucella abortus
C:Species: Brucella abortus
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 26-Aug-1999
C:Accession: S22347
R:Gor, D.; Mayfield, J.E.
Biochim. Biophys. Acta 1130, 120-122, 1992
A:Title: Cloning and nucleotide sequence of the Brucella abortus groE operon.
A:Reference number: S22346; MUID:92182006; PMID:1347461
A:Accession: S22347
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-546 <GOR>
A:Cross-references: EMBL:M82975; NID:g144109; PIDN:AAA22997.1; PID:g144111
C:Superfamily: chaperonin groEL

Query Match 67.3%; Score 1823; DB 2; Length 546;
Best Local Similarity 68.9%; Pred. No. 2.6e-81;
Matches 378; Conservative 64; Mismatches 99; Indels 8; Gaps 5;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLGPKGRNVIEKSPGSPVITKDGVSVAKEI 60
DB 1 MAARKVKFGRTAREKMLRGVDILADAVKVTGLGPKGRNVIEKSPGSPVITKDGVSVAKEV 60
QY 61 ELEDKFNMGAGQVKEVAPKTSIDTAGDGTATTATVLAQAIYREGVKLVAAGRNPMIAIKRGI 120
DB 61 ELEDKFNMGAGQMLREVASKNTNDTAGDGTATTATVLAQAIYREGVKLVAAGRNPMIAIKRGI 120
QY 121 DKAVVAVTKELSDITKTPRQKEIAQVGTTSANSDDTIGNIIIAEAMAKVKGGVITVEEA 180
DB 121 DLAVNEVVAELLKKAKKINTSEEAQVGTISANAERIGKMAEAMQKVGNEGVITVEEA 180
QY 181 KGLTTLDVVEGKDFRGYLSPYFVTNPEKMWCELDNPIYILCNEKKITSMKMDLPILAEV 240
DB 181 KTAETELVEVGEQMDRGYLSPYFVTNPEKMWCELDNPIYILCNEKKITSMKMDLPILAEV 240
QY 241 AKVNRPLIIIAEDVEGEALATLVNKLRGALQVAVKAPGGERRKAMLEDAIILTGEA 300
DB 241 VQTSKPLIIIAEDVEGEALATLVNKLRGALQVAVKAPGGERRKAMLEDAIILTGTG 300
QY 301 IFEDRGKLENVSSISLGTAKRVVDKENTTTVDGAGKSEDIKARVKQIRAQIBETSSDY 360
DB 301 ISEDLGKLESVTLMDLGRKKVSIKENTTTVDGAGKAEIDARVQGIKQIIBETSSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAEEGIVPGGGTAFV 420

361 DREKLOERLAKLVGGVAVIRVGGATEVEVEKKEKRDVDDALNATRAAEEGIVAGGSTALL 420
421 R-SIKVLDDIKPADDDELAGLNIIRSLRQIAANAGYEGSIVVEKRVKDFGFGNA 478
421 RASTKI--TAKGVNADQEAQINIVRAIQAPARQITTNAGEEASVIVGKILENTSETFGY 478
479 NAASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEPKDMPPG--G 536
479 NTANGEYGDLLISLGIQVDPKVTALQNAASVAGLLITTEAMIAEL--PKDAAPAGMPG 536
537 GGMGMDGM 545
537 GGMGMDGM 545

Search completed: January 28, 2004, 13:17:49
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:10:35 ; Search time 17 Seconds

(without alignments)

1515.920 Million cell updates/sec

Title: US-09-077-574a-2

Perfect score: 2708

Sequence: 1 MASKELFDKAREKLSRGV.....KMPMPGGGGMGGMGMDGMY 548

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1891.5	69.8	547	1	CH60_RALSO
2	1880	69.4	546	1	CH60_ACRAC
3	1872.5	69.1	546	1	CH60_BURCE
4	1868.5	69.0	546	1	CH60_BURPS
5	1868.5	69.0	546	1	CH60_BURVI
6	1868	69.0	546	1	CH60_RHOPA
7	1867.5	69.0	546	1	CH60_BURTH
8	1867.5	69.0	547	1	CH60_BURHE
9	1867	68.9	545	1	CH61_RHIME
10	1866	68.9	551	1	CH60_AMOPS
11	1864.5	68.9	547	1	CH60_BARQU
12	1864	68.8	545	1	CH60_RHIME
13	1862.5	68.8	547	1	CH60_XYLFA
14	1852.5	68.4	540	1	CH60_RHOMR
15	1852.5	68.4	546	1	CH60_CHRVI
16	1847	68.2	546	1	CH60_XANAC
17	1845	68.1	545	1	CH60_RHOCA
18	1844.5	68.1	545	1	CH60_RHTLV
19	1844	68.1	546	1	CH60_XANCH
20	1843.5	68.1	547	1	CH60_BORPE
21	1843.5	68.1	552	1	CH60_COXBU
22	1841	68.0	546	1	CH60_XANCP
23	1840.5	68.0	542	1	CH65_RHIME
24	1840.5	68.0	544	1	CH60_AGR5
25	1834.5	67.7	546	1	CH61_RHOSH
26	1834.5	67.7	552	1	CH60_PSEST
27	1834	67.7	542	1	CH62_RHILE
28	1833.5	67.7	539	1	CH61_BRAJA
29	1832.5	67.7	545	1	CH60_PARDE
30	1828.5	67.5	545	1	CH60_CHLTE
31	1826	67.4	542	1	CH62_RHILO
32	1825.5	67.4	547	1	CH60_PSEAE
33	1823	67.3	546	1	CH60_BRUAB

Q8fx87 brucella su
P35861 bradyrhizob
P35470 rhizobium m
P57006 neisseria m
Q8yb53 brucella me
Q98394 rhizobium l
Q98ax9 rhizobium l
P42385 neisseria m
P35635 bartonella
Q9xau7 alteromonas
Q981j9 rhizobium l
P29842 neisseria g

34 1819 67.2 546 1 CH60_BRUSU
35 1818 67.1 549 1 CH62_BRAJA
36 1816 67.1 542 1 CH62_RHIME
37 1814.5 67.0 544 1 CH60_NEIMA
38 1814 67.0 546 1 CH60_BRUME
39 1812.5 66.9 551 1 CH64_RHILO
40 1810 66.8 552 1 CH63_RHILO
41 1808.5 66.8 544 1 CH60_NEIMB
42 1805.5 66.7 543 1 CH60_BARBA
43 1804.5 66.6 547 1 CH60_ALTHA
44 1803.5 66.6 549 1 CH65_RHILO
45 1802.5 66.6 544 1 CH60_NEIGO

ALIGNMENTS

RESULT 1

CH60_RALSO STANDARD; PRT; 547 AA.
AC Q8YLP8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL OR MOFA OR RSC0642 OR RSO1546.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
Gaspin C., Javie M., Moisan A., Robert C., Saurin W., Schiex T.,
Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002)
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL646060; CAD14172.1; -
DR HAMAP; MF 00600; -; 1
DR InterPro; IPR001844; Chaperin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 547 AA; 57404 MW; 0487E4650867C25A CRC64;

Query Match

Best Local Similarity 69.8%; Score 1891.5; DB 1; Length 547;

Matches 380; Conservative 70; Mismatches 96; Indels 1; Gaps 1;

1 MASKEILFDKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSGFSPVITKDCVSVAKEL 60

db
1 MAAKDVFGDAAAKMVEGVNLANAVKVTLGPKGRNVILERSFGGPTVTKDGVSAKEI 60

61	ELEDKFENNCAQMVVEVAPKTS	DIAGD	TTTATVLAQAIYREGVKLVAAAGRN	PMATKRG	120
QY		:		:	
61	ELEDKFENNCAQMVVEVAPKTS	DIAGD	TTTATVLAQAIYREGVKLVAAAGRN	PMATKRG	120
		:		:	

db 61 ELKDKLQNMGAQMVKEVASTSDNAGDGTITATVLAQSLVREGMKIVAAAGNMFMELARGI 129

QY
121 DKAAVAVTKELSDITKPTRDQKEIAQVGTSANSDDTLIGNLLAEAWAKVRGGVVILVEER 100

Db 121 DRAVAAAEEELKKISRTTISNELAQVGAISANSDSIGARHAEADKAGADGVII1250 1250

[illegible]

DB 181 KSELELDVV EGMQF DRGILGFIITAH ENQ V QZKALIVZETP
241 *VWDBBIITATFVGEFAIATIWNKURGALOVVAVKAPFGERRKAMLEDIAITGCEA 300

Z41 AKVNFLELLIYADSVGGHAFILFALRERCTK
Z41 AKGRPIITVFADFVEGALATLVNNIRGLIKTAVKAPGFDRRKAMLEDIALTGGQV 300

DD 211 REPORT OF PERSONNEL

OV 301 IFEDRGKLENVSLSSLGTAKEVWIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSD 360
+ + + + +
+ + + + +

301 IAEVGLTLEKATLNDIGQAKRVEIGKENTIIDGAGDARNEARVQVRAQLEATSDY 360

361 DREKLERLAKLVGGVAIVHVGAAATETEMKEKKORVEDALNATRAAAVEEGIVPCGGTAFV 420

361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGVALL 420

421 RSIKVLDDIKPADDDLAGLNITRSLEEFRLQIAANAGYEGSIVKEVREPDKGFNA 480

421 PARALISGLKANADQAGIKIVIRAMEEPLRQIVTNAGDEASVVVAVIACKNGYGYNA 480

QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEPKDMPGGMG 540

Db 481 STGEYGLVEMGVLDPTKVTRTALQNAASVASLMLTTDCAVELPKDDAAPAMP-GGMGG 539

QY 541 MGGMDGM 547

540 MCGMDGM 546

DEBIT 2

RECORD #	CH60 ACEAC	ID_CH60 ACEAC	STANDARD;	PRT;	546 AA.
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					
87					
88					
89					
90					
91					
92					
93					
94					
95					
96					
97					
98					
99					
100					

AC Q8GBD2;
DT 15-SEP-2003 (Rel. 42, Created)

DT	15-SEP-2003 (Rel. 42, Last sequence update)
DT	15-SEP-2003 (Rel. 42, Last annotation update)

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL.

OS. Acetobacter aceti.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;

OC Acetobacteraceae; Acetobacter.
OX NCBI_TaxID=435;
OX

RN [1]
RP SEQUENCE FROM N.A.
RC CRYSTALLINITY 2203.

RC SIRAIN=110 3203,
RA Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,
RA Venacida F.

RT "Cloning and Characterization of groEL Operon in Acetobacter aceti." ;
PL J. Biosci. Bioeng. 94:140-147 (2002).

- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress

CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC (transmembrane) family.

CC -|- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -|- SIMILARITY: It is produced through a collaboration

CC This SWISS-PROT entry is copyright © 1993 by the Swiss Institute of Bioinformatics. There are no restrictions on its use.

CC the European Bioinformatics Institute. There are no restrictions on the use of the data by non-profit institutions as long as its content is in no way

"Nucleotide sequence comparison of the groE operon of Burkholderia spp.";

RT submitt (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF104907; AAC79087.1; --

CC HSPSP; P06139; 1GRL.

CC HAMAP; MF 00600; -; 1

CC InterPro; IPR001844; Chaperonin Cpn60.

CC InterPro; IPR002423; Cpn60/TCP-1.

CC Pfam; PF00118; cpn60 TCP1; 1.

CC PRINTS; PR00298; CHAPERONIN60.

CC PRINTS; PR00304; TCOMPLEXTCP1.

CC PROSITE; PS00296; CHAPERONIN60; 1.

CC Chaperone; ATP-binding.

CC SEQFEATURES 546 AA; 56980 MW; 174B9934345E7315 CRC64;

Query Match 69.1%; Score 1872.5; DB 1; Length 546;

Best Local Similarity 68.4%; Pred. No. 1.6e-82;

Matches 374; Conservative 75; Mismatches 95; Indels 3; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLPGKRNVLIEKSFQSVLTGKGVSAKEI 60

Db 1 MAADKVVFGDSARSKVVGVNLANAVKVTGLPGKRNVLERSFGFTVTGKGVSAKEI 60

QY 61 ELEDKFNAGQVMKEVAPKTSIDAGDGTATVLAQAIVREGVKLVAAGRNPMKKEGI 120

Db 61 ELKDKLQNNGAQVMKEVAVAKTSIDNAGDGTATVLAQSVREGKLVASGNPMKKEGI 120

QY 121 DRAVAVATKELSDITKPTDQKEIAQVGTISANSDTTIGNIIIAEMAKVKGGVITVEEA 180

Db 121 DRAVAAVBEELKKISKPTCTNKEIAQVGSISANSDDSGIDRIAEMDKVKGEGVITVEDG 180

QY 181 KGLETTLDVVEGKMFDRGYLSPFVFNPEKMYCELDNPFVILCNEKLTSMKMDMLPILQV 240

Db 181 KSLADELDVVEGQFDRGYLSPFFNNPDKQVAVLDNPFVLLHDKKVSNIREDLLPVLQV 240

QY 241 AKVNPLLIIDVDEGEALATVIVNKLQALQVAVKAPFGGERKAMLEIDAILTGQV 300

Db 241 AKAGRPLLIADVDEGEALATLVANNIRGLTKVAVKAPFGDREKAMLEIDAILTGQV 300

QY 301 IFEDSGIKLVNLSLSLGTAKRVVIDKENTTIVDGAGKSDIKARVKQIRAIQIETSDDY 360

Db 301 IAEETGLTLEKATLAEGLQAKREIVGKENTTIIDGAGEAASIEARVKQVRAQIEEATSDY 360

QY 361 DREKQLERLAKLVGVAVTHVGAATETEMKCKKQVEDALNATRAAEEGVPGGGTAFV 420

Db 361 DREKQLERQVAKLGGVAVIKVGAATEVEMKCKKARVEDALHATRAAEEGVVAGGVGVALI 420

QY 421 RSIKVLDDIKPADDELALAGNIIRSLBEPLRQIAANAGYEGSVIWEKREPDKQFGFNA 480

Db 421 RARTALAGLTGANADQAGIKVILPAMEEPLRQIVTNGGEEASVVAAVAGKNGYGYNA 480

QY 481 ASGEYEDLIKAGVIDPKVKTRIALQNAASVALLITTECAIAEKPEPKKMPMPGGMGWG 540

Db 481 ATGEYVDVMEAGVVDPTKVTTRTALQNAASVAGLLITDAVAEL--PKEADPMPGGMPGG 538

QY 541 MGGMDGM 547

|||||

Db 1 MAADVVFGDSARSQKVEGVNLANAVKVTGLGKRNVLERSFGPTVTKDGSVAKI 60
Qy 61 ELEKFEKNGAQMKVEKAPKTSIAGDCTTATVLAQAIYREGVKLVAAAGRNPMKRG 120
Db 61 ELKOKLQNGAQMKVEKASKISDNAGDGTATVLAQSIYREGMKYVAGSMNPMKRG 120
Qy 121 DKAVAVTKELSDITKPTRDQKEIAQVGTISANSSTTTIGNIIAEMAKVKGCVITVEA 180
Db 121 DKAVAAVEELKISKPTTNTKEIAQVGSISANSSSIGDRIAEAMDVKVGEVITVEDG 180
Qy 181 KGLTLLDVVEGKFDGKYLSPYFVTPNPEKMKVCELDNPNVILCNEKKITSMKMDLPLEOV 240
Db 181 KSLADELDVVEGQMDRGYLSFYFNNPKQVAVLDNPPVLLHDKKVSINRDLPLVEQV 240
Qy 241 AKVNEPLLIIAEDVGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLGGEA 300
Db 241 AKAGPLLIIAEDVGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLGGQV 300
Qy 301 IFEDRGIKLENVSLSLGTAKVWIDKENTTTVDGAGKSEDIKARVQTRAQIEETSSDY 360
Db 301 IAEETGLTLEKATLAELGOAKRIEVEKENTTTIDGAGEAASIEARVQVTRQIEEATSDY 360
Qy 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAVERGIVPGGTAFFV 420
Db 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAVERGIVAGGVALI 420
Qy 421 RSIKVLDITKPADDDDELGLNIIRSLERPLRQIAANAGYEGSVVVEKVPKDGFGFNA 480
Db 421 RARTAIAGLTGANADQAGIKVIRAMEPLRQIVTNGGEEASVVVAAGAAGKNGYNA 480
Qy 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEPKDMPGGMGG 540
Db 481 ATGEYVDMVEAGVDPKVTTRTALQNAASVAGLLTTDAVAEL--PKEDAPMPGMPGG 538
Qy 541 MCGMDGM 547
Db 539 MCGM-GM 544

RESULT 6
CH60 RHOPA STANDARD; PRT; 546 AA.
ID -CH60 RHOPA
AC Q93MH1;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL.
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiales; Rhodopseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao M., Zhu C.R., Qian X.M., Zheng P., Chen Y.Y.;
RT "Cloning and Sequencing of the groEL Operon of Rhodopseudomonas
palustris";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).
EMBL; AF104908; AAC79089.1; -
HSSP; P06139; 1GRL.
DR HAWAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaperonin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
DR Chaperone; ATP-binding.
SQ SEQUENCE 546 AA; 57010 MW; 47FA893959221218 CRC64;

Query Match 69.0%; Score 1868.5; DB 1; Length 546;
Best Local Similarity 58.2%; Pred. No. 2.5e-82;
Matches 373; Conservative 75; Mismatches 96; Indels 3; Gaps 2;
1 MASKELLFDKAREKLSRGVDKLANAVKVTGLGKRNVLERSFGSPVITKDGVSVAKEI 60

Qy 301 IAEETGLTLEKATLAELGOAKRIEVEKENTTTIDGAGEAASIEARVQVTRQIEEATSDY 360
Qy 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAVERGIVPGGTAFFV 420
Db 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKORVEDALNATRAAVERGIVAGGVALI 420
Qy 421 RSIKVLDITKPADDDDELGLNIIRSLERPLRQIAANAGYEGSVVVEKVPKDGFGFNA 480
Db 421 RARTAIAGLTGANADQAGIKVIRAMEPLRQIVTNGGEEASVVVAAGAAGKNGYNA 480
Qy 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLLTTECAIAEKPEPKDMPGGMGG 540
Db 481 ATGEYVDMVEAGVDPKVTTRTALQNAASVAGLLTTDAVAEL--PKEDAPMPGMPGG 538
Qy 541 MCGMDGM 547
Db 539 MCGM-GM 544

RESULT 5
CH60 BURVI STANDARD; PRT; 546 AA.
ID -CH60 BURVI
AC Q9ZFD8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL.
OS Burkholderia vietnamiensis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=60552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 11319;
RA Zysk G., Spietstoesser W.D., Neubauer H.;
RT "Nucleotide sequence comparison of the groE operon of Burkholderia
sp.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).
EMBL; AF104908; AAC79089.1; -
HSSP; P06139; 1GRL.
DR HAWAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaperonin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
DR Chaperone; ATP-binding.
SQ SEQUENCE 546 AA; 57010 MW; 47FA893959221218 CRC64;

Query Match 69.0%; Score 1868.5; DB 1; Length 546;
Best Local Similarity 58.2%; Pred. No. 2.5e-82;
Matches 373; Conservative 75; Mismatches 96; Indels 3; Gaps 2;
1 MASKELLFDKAREKLSRGVDKLANAVKVTGLGKRNVLERSFGSPVITKDGVSVAKEI 60

or send an email to license@isb-sib.ch.

```

CC EMBL; AF406639; AAK94943.1; -.
CC HAMAP; MF_00600; -.
CC InterPro; IPR001844; Chaprinin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC Pfam; PF00118; cpn60_TCF1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PROSITE; PS00304; TCOMPLEXTCF1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC SEQUENCE 546 AA; 57937 MW; 163C7115D68FF8 CRC64;

Query Match          69.0%; Score 1868; DB 1; Length 546;
Best Local Similarity 68.2%; Pred. No. 2.7e-82;
Matches 374; Conservative 75; Mismatches 95; Indels 4; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
Db 1 MAADKVFDTDARDMLRGVNLDAQVKTLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
QY 61 ELEDKFENMGAGQVKEVAPKTSIDAGDGTATVLAQALYREGVKLVAAGRNPMAIKRGI 120
Db 61 ELSDFENMGAGQVKEVARSINDEAGDGTATVLAQALYREGVKLVAAGRNPMDLKRGI 120
QY 121 DKAVAVTKELSDITKPTRDQKEIAQVGTISANSPTTIGNIIAEMAKVKGKGVITVEEA 180
Db 121 DLATAKVVESIKAAASRPVNDQHEVAQVGTISANGEAQIGRFIADAMQVKGEGVITVEEN 180
QY 181 KGLETTLLDVVEGKFDGRLSPYFVTNPEKVMCELDNPIYILCNEKKITSMKMDLPILQV 240
Db 181 KGLETEVEVEGKFDGRLSPYFVTNADKNMTAELEDVFIHHEKLSLQPMVPLESV 240
QY 241 AKVNRPLLIITAEVDEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAIITGGEA 300
Db 241 IQAQRPLLIITAEVDEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAIITGGEA 300
QY 301 IFEDRGKLENVSLSLGTAKRVVIDKENTTIVDGAKESEDIKARVKQIRAOIEETSSDY 360
Db 301 ISEDLGKLENVTIDMLGRKVSINKDNTTIVDGAKESEDIKARVKQIRAOIEETSSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATEEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLOERLAKLVGGVAVIRVGGTVEVKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
QY 421 RSIKVLDDIKPADDELALGNIIRSLRLEPLRQIAANAGVEGTVVEKVRPKD-GFGFN 479
Db 421 QAGKVLDDLTGENPDQAGTIVRRALEAPLRQIAQAGVGVAGKVRRESDDKAFGN 480
QY 480 AASGEYEDLIKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPPKMPGGMG 539
Db 481 AQTEYEDGMKFGVIDPAKVRTALEDAASVASILLITTEAVIAADKPEPKS---APAGGMG 537
QY 540 GMGMDGM 547
Db 538 GMGMDGM 545

```

RESULT 7

```

CH60_BURTH          STANDARD;          PRT;          546 AA.
AC P58723;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 60 kDa chaperonin (Protein Cpn60) (update)
GN GROEL OR GROEL.
OS Burkholderia thailandensis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Burkholderiaceae; Burkholderia.
RN [1]_TaxID=57975;
RP SEQUENCE FROM N.A.

```

RA Woo P.C.Y., Woo G.K.S., Yuen K.Y.;
 RT "Burkholderia thailandensis chaperonin GroEL gene, complete
 sequence."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
 proper assembly of unfolded polypeptides generated under stress
 conditions (By similarity).
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 7 subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

```

CC EMBL; AF454383; AAL49762.1; -.
CC HAMAP; MF_00600; -.
CC InterPro; IPR001844; Chaprinin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC Pfam; PF00118; cpn60_TCF1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCF1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC SEQUENCE 546 AA; 57145 MW; 913CD01A5C131A06 CRC64;

```

Query Match 69.0%; Score 1867.5; DB 1; Length 546;
 Best Local Similarity 68.2%; Pred. No. 2.8e-82;
 Matches 373; Conservative 76; Mismatches 95; Indels 3; Gaps 2;

```

QY 1 MASKEILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
Db 1 MAADKVFDTDARDMLRGVNLDAQVKTLGPKGRNVVIEKSPGVITKDGVSVAKEI 60
QY 61 ELEDKFENMGAGQVKEVAPKTSIDAGDGTATVLAQALYREGVKLVAAGRNPMAIKRGI 120
Db 61 ELKDKLQNMGAQVKEVAKTSIDAGDGTATVLAQALYREGVKLVAAGRNPMDLKRGI 120
QY 121 DKAVAVTKELSDITKPTRDQKEIAQVGTISANSPTTIGNIIAEMAKVKGKGVITVEEA 180
Db 121 DKAVAAVEELKKISKPTTKEIAQVGSISANSPTTIGNIIAEMAKVKGKGVITVEEG 180
QY 181 KGLETTLLDVVEGKFDGRLSPYFVTNPEKVMCELDNPIYILCNEKKITSMKMDLPILQV 240
Db 181 KSLADELDVVEGKFDGRLSPYFVTNPEKVMCELDNPIYILCNEKKITSMKMDLPILQV 240
QY 241 AKVNRPLLIITAEVDEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAIITGGEA 300
Db 241 AKAGRPLLIITAEVDEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAIITGGEA 300
QY 301 IFEDRGKLENVSLSLGTAKRVVIDKENTTIVDGAKESEDIKARVKQIRAOIEETSSDY 360
Db 301 IABETGLTEKATLAEGLQAKRLEVKNITIDGAGEAVNIEARVKQIRAOIEETSSDY 360
QY 361 DREKLOERLAKLVGGVAVIHVGAATEEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
Db 361 DREKLOERLAKLVGGVAVIKVGAATEEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
QY 421 RSIKVLDDIKPADDELALGNIIRSLRLEPLRQIAANAGVEGTVVEKVRPKD-GFGFN 480
Db 421 RARTAAALTGVAQNAQAGIKVLRAMEEPLRQIVTNGGEEASVVAAGKNGTGYNA 480
QY 481 AASGEYEDLIKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPPKMPGGMG 540
Db 481 ATGEYEDMVAGVDPKVTIRIALQNAASVASILLITTECAIAEKPPKMPGGMG 540
QY 541 GMGMDGM 547

```

539 MGGM-GM 544

Matches 385; Conservative 61; Mismatches 96; Indels 7; Gaps 4;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 DB 1 MAAKEVFGREARERLLRGVDILANAVKVTLPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 QY 61 ELEDKFENMGQAVKVPKTSIAGDGTATVLAQAIYEGYKLVAAAGNPWAIKEGI 120
 DB 61 ELEDKFENMGQAVKVPKTSIAGDGTATVLAQAIYEGYKLVAAAGNPWAIKEGI 120
 QY 121 DKAVVATKELSDITKPTDRDQKEIAOYGTISANSDDTTIGNIIAEMAKVKGKGVITVEEA 180
 DB 121 DAAVDEVVANLFKAKKIQTSAEIAQVGTISANGAAELCKMIADAMEKVGNEGVITVEEA 180
 QY 181 KGLTTLDDVVEGMKFDRCGLSPYFVTNPKEKVCELDNPYILCNEKKTSMKMDMLPILQV 240
 DB 181 KTAETELVVEGMQFDRGLSPYFVTNAEKWVADLDDPVLITHEKKLSNLQSLLPVLEAV 240
 QY 241 AKVNRPLIIADVEGEALATLVNKLKGLKIAAVKAPGFGERRKAMLEDAITLTGGEA 300
 DB 241 VQSGKPLIIADVEGEALATLVNKLKGLKIAAVKAPGFGERRKAMLEDAITLTGGEA 300
 QY 301 IREDGKILENVSLSLGTAKRVVIDKENTTIIVDAGKSEDIKARVQIRAOIETSSDY 360
 DB 301 ISEDVGKILENVSLSLGTAKRVVIDKENTTIIVDAGKSEDIKARVQIRAOIETSSDY 360
 QY 361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKORVEDALNATRAAVEGIVPGGTAFAV 420
 DB 361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKORVEDALNATRAAVEGIVPGGTAFAV 420
 QY 421 RSIKVLDDITKPADDDDELGINIIRSLERPLQIAANAGVSGIVVEKVRPEK-DGFGFN 479
 DB 421 RAANAL-TVKGSNPDOEAGINIVRRALQAPARQIATNAGEEAAIIVGKVLNNADTFGYN 479
 QY 480 AASGEYEDLIKAGVIDPKKVTTRIALONAAVASLITTECAIAEKPEPKDM---PMPGG 536
 DB 480 TATGEGDILALGIVDPKVVRSALQNAASIASLLITTEAMVAE--VPKDTVPVMPGG 537
 QY 537 GMGGMGMD 545
 DB 538 GMGGMGMD 546

RESULT 9

CH61 RHIME STANDARD; PRT; 545 AA.

ID CH61 RHIME STANDARD; PRT; 545 AA.

AC P35469; (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1).

GN GROEL1 OR GROEL1 OR GROEL1 OR R00792 OR SMC00913.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

OC NCBI TaxID=382;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=93231539; PubMed=8097179;

RA Rusanwanwa E., Gupta R.S.;

RT "Cloning and characterization of multiple groEL chaperonin-encoding

RT genes in Rhizobium meliloti."

RL Gene 126:67-75(1993).

EN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=1021;

RA Ogawa J.;

RT Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

EN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RX

DB 539 MGGM-GM 544

RESULT 8

ID CH60 BARHE STANDARD; PRT; 547 AA.

AC O33963; 087267;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein

DE 60).

GN GROEL OR GROEL OR MOXA.

OS Bartonella henselae (Rochalimaea henselae).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bartonellaceae; Bartonella.

OC NCBI TaxID=38323;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 49882 / Houston 1;

RX MEDLINE=97419519; PubMed=9274034;

RA Haake D.A., Summers T.A., McCoy A.M., Schwartzman W.;

RT "Heat shock response and groEL sequence of Bartonella henselae and

RT Bartonella quintana."

RL Microbiology 143:2807-2815(1997).

EN [2]

RP SEQUENCE OF 1-543 FROM N.A.

RC STRAIN=ATCC 49882 / Houston 1;

RA Marston E.L., Sumner J.W., Regnery R.L.;

RT "Evaluation of intraspecies genetic variation within the 60 kDa heat

RT shock protein (groEL) gene of Bartonella species: a new phylogenetic

RT analysis tool."

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

EN [3]

RP SEQUENCE OF 1-408 FROM N.A.

RC STRAIN=ATCC 49882 / Houston 1;

RX MEDLINE=97373904; PubMed=9230387;

RA Sumner J.W., Nicholson W.L., Maesung R.F.;

RT "PCR amplification and comparison of nucleotide sequences from the

RT groEL heat shock operon of Ehrlichia species."

RL J. Clin. Microbiol. 35:2087-2092(1997).

CC -1- FUNCTION: Prevents misfolding and promotes the refolding and

CC proper assembly of unfolded polypeptides generated under stress

CC conditions (By similarity).

CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

CC 7 subunits (By similarity).

CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

EMBL; U78514; AAB69094.1; -

EMBL; AF014829; AAD04238.1; ALT_INIT.

EMBL; U96734; AAB65637.1; -

HSSP; P06139; 1GRL.

HMAP; MF 00600; -; 1.

InterPro; IPR001844; Chaprinin_Cpn60.

InterPro; IPR002423; Cpn60/TCP-1.

Pfam; PF00118; Cpn60_TCP1; 1.

PRINTS; PR00298; CHAPERONIN60.

PRINTS; PR00304; TCOMPLEXICP1.

PROSITE; PS00296; CHAPERONIN_CPN60; 1.

KW Chaperone; ATP-binding; Heat shock.

SEQUENCE 547 AA; 57625 MW; EDE25D566D5B10D2 CRC64;

Query Match 69.0%; Score 1867.5; DB 1; Length 547;

Best Local Similarity 70.1%; Pred. No. 2.8e-82;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sincrozobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94192; AAA26285.1; -;
CC EMBL; U19726; AAA61955.1; -;
CC EMBL; AL591785; CAC45364.1; -;
CC DR PIR; JN0509; JN0509.
CC DR HSP; P06139; 1GRL.
CC DR HAMAP; MF_00600; -; 1.
CC DR InterPro; IPR001844; Chaperonin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCF-1.
CC DR Pfam; PF00118; cpn60_TCF1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCPI.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC DR PROSITE; PS00304; TCOMPLEXTCPI.
CC KW Chaperone; ATP-binding; Heat shock; Multigene family;
CC Complete proteome.
CC FT CONFLICT 12 12 A -> G (IN REF. 1).
CC SQ SEQUENCE 545 AA; 57687 MW; 4CEP358957ELB45A CRC64;
CC -----
CC Query Match 68.9%; Score 1867; DB 1; Length 545;
CC Best Local Similarity 70.7%; Pred. No. 3e-82;
CC Matches 388; Conservative 58; Mismatches 97; Indels 6; Gaps 4;
CC -----
CC QY 1 MASKEILFDKAREKLSRGYDKLANAVKTLGPKGRNVIEKFGSPVITKDGVSVAKEI 60
CC Db 1 MAAKEVKGFSAREKMLRGVDILADAVKVTGLGPKGRNVVDKSGFAPRITKDGVSVAKEI 60
CC -----
CC QY 61 ELEDKFNMGQVMVKEVAPKTSIDAGDGTATVLAQAIYREGVGLVAAGRNPMIAIKRGI 120
CC Db 61 ELEDKFNMGQVMREVASKINDIAGDGTATVLAQAIYREGAKVAAGRNPMIDIKRGI 120
CC -----
CC QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEAMAKVGKGVIVVEEA 180
CC Db 121 DLAVAEVWKDLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQVGNQGVIVVEEA 180
CC -----
CC QY 181 KGLFTLDVVEGKFKRGYLSPFVFNPEKWCCELNDNYILCNKKITSMKMLPILEQV 240
CC Db 181 KTAETELEVVEGQFDRGYLSPFVFNPEKRWADLEDAFILLHEKKLSNLQAMLPVLEAV 240
CC -----
CC QY 241 AKVNRPLLIIEAEVEGEALATLVNKLRGALQVAVKAPGFGERRKAMLEDTAILTGGEA 300
CC Db 241 VQTGKPLLIIEAEVEGEALATLVNKLRLGKLTAAVAPGFGRRKAMLEDTAILTGGV 300
CC -----
CC QY 301 IFEDRGKILEVSSLSGTAQRVVDKENTTIVDGAGKSGEDIKARVKQIRQAIEETSSDY 360
CC Db 301 ISEDLGKILESVDLMLGRKKVSIKENTTIVDGAGKSDIEGRVAQIAKIEETTSY 360
CC -----
CC QY 361 DREKLOERLAKLVGGVAVIHVGAATEEMKEKORVEDALNATRAAVERGIVGGGTAFV 420
CC Db 361 DREKLOERLAKLVGGVAVIRVGATEVEVEKEDKRDIDALNATRAAVERGIVGGGVALL 420

QY 421 R-SIKVLDDIKPADDELALAGLNIRSRLEELPRLQRTAANAGYEGSIIVVKVRPK-DGRGF 478
Db 421 RSSVKI--TVRGENDDDQAGVNIIVRRALQSPARQIVENAGDEASIVVGKILEKNTDDFGY 478
QY 479 NAASGEYEDLLKAGVIDPKKYTRIALQNAASVASILLTTECAIAEKPEPKKDMPPGGGM 538
Db 479 NAQTGEYDMLTAMGLIDPKVVRALQDAASVASILLTTEAMIAEL--PKDAPAMPGGM 536
QY 539 GGMGGMDGM 547
Db 537 GGMGGMDMM 545

RESULT 10
CH60_AMOPS
ID CH60_AMOPS STANDARD; PRT; 551 AA.
AC P26004;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR MOPA.
OS Amoeba proteus symbiotic bacterium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae.
OX NCBI_TaxID=2728;
RN [1]
RP SEQUENCE FROM N.A.
RA Ahn T.I., Leeu H.K., Kwak I.H., Jeon K.W.;
RT "Nucleotide sequence and temperature-dependent expression of groEL
RT gene isolated from symbiotic bacteria of Amoeba proteus.";
RL Endocyt. Cell Res. 8:33-44(1991).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M86549; AAC09381.1; -;
CC DR PIR; JC2562; JC2562.
CC DR HSP; P06139; 1GRL.
CC DR HAMAP; MF_00600; -; 1.
CC DR InterPro; IPR001844; Chaperonin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCF-1.
CC DR Pfam; PF00118; cpn60_TCF1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCPI.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC SQ SEQUENCE 551 AA; 58332 MW; A25895CBF9585DA0 CRC64;
CC -----
CC Query Match 68.9%; Score 1866; DB 1; Length 551;
CC Best Local Similarity 67.9%; Pred. No. 3.4e-82;
CC Matches 373; Conservative 74; Mismatches 96; Indels 6; Gaps 2;
CC -----
CC QY 3 SKELLFDKAREKLSRGYDKLANAVKTLGPKGRNVIEKFGSPVITKDGVSVAKEI 62
CC Db 2 AKELRFGDARQOQMLAGVNLADRVKATMGPSGRNVVLSRFGAPTIVTKDGVSVAKEIF 61
CC -----
CC QY 63 EDKFNMGQVMVKEVAPKTSIDAGDGTATVLAQAIYREGVGLVAAGRNPMIAIKRGI 122
CC Db 62 ENRFRNMGQVMVKEVAPKTSIDAGDGTATVLAQAIYREGVGLVAAGRNPMIAIKRGI 121

123 AVAVTVKELSDITKPTDRDQKIAQVGTISANSDDTTIGNIIAEAMAKVGGVITVERAKG 182
122 AVTAITKELQVSKPCDKGKAIAQVGTISANSDDQIGSIIIAEAMKEVKGEGVITVEDGNG 181
183 LETTLDVVEGKFDGKGLSPYFVTNPEKMWCELDNPYILCNEKKITSMKMDLPFILEQVAK 242
182 LENELSVVEGQFDRGYISPIFINNOQNSAELEHPPFILLVDKTIATIRDMLSVLEAVAK 241
243 VNRPLIIIAEDVEGEALATLVNKLRGALVAVKAPGFGERRKAMLEIDAILTGGEAIF 302
242 SGRPLIIIAEDVEGEALATLVNKLRGALVAVKAPGFGERRKAMLEIDAILTGGEAIF 301
303 EDRIKLENSVLSLGTAKRVVDKENTIVDQKSEDIKARVQIRAIQAEETSSDYDR 362
302 BEIGTSLETSLESIAGTAKRVVDKENTIVDQKSEDIKARVQIRAIQAEETSSDYDR 361
363 EKLQERLAKVGGVAVIHVGAATEMEKKDVEDALNATRAAEEGIVPGGGTAFVRS 422
362 EKLQERLAKVGGVAVIHVGAATEMEKKDVEDALNATRAAEEGIVPGGGTAFVRS 421
423 IKVLDDIKPADDDDELALNIIIRRSLEPLRQIAANAGVEGSIIVKVRPKDGFENAS 482
422 QKVLGKGNADQDMGINILRRATIESPLRQIVANAGYESSIVNKNVAEHKDNFGFNAAT 481
483 GEYEDLITKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPEPKDMPG---GGM 538
482 GOYGDVMEVGIIDPTKVTIRIALQNAASVASILLITTECAIAEKPEPKDMPG---GGM 539
539 GGMGGMGMD 547
540 GGMGGMGMD 548

RESULT 11

CH60_BAROU STANDARD; PRT; 547 AA.
ID CH60_BAROU STANDARD; PRT; 547 AA.
AC O33964;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein 60).
DE GROEL OR GROEL OR MOPA.
GN Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51694;
RX MEDLINE=97419519; PubMed=9274034;
RA Haake D.A., Summers T.A., McCoy A.M., Schwartzman W.;
RT "Heat shock response and groEL sequence of Bartonella henselae and Bartonella quintana".
RL Microbiology 143:2807-2815 (1997).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U78515; AAB69095.1; -

DR HSP; P06139; IREL.
DR HAMAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Heat shock.
KW SEQUENCE 547 AA; 57610 MW; 8620CFC50BF2D6F4 CRC64;
Query Match 68.9%; Score 1864.5; DB 1; Length 547;
Best Local Similarity 70.1%; Pred. No. 4e-82; Mismatches 96; Indels 7; Gaps 4;
Matches 385; Conservative 61;
QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTLPKGRNVIKSGSPVITKDGVSVAKEI 60
DB 1 MAKEKFKGREAREKLLRGVDILANAVKVTLPKGRNVIDKSGFAPRITKDGVSVAKEI 60
QY 61 ELEDKFNMGAGQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAGRNPMALKRI 120
DB 61 ELEDKFNMGAGQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAGRNPMALKRI 120
QY 121 DKAVVAVTKELSDITKPTDRDQKIAQVGTISANSDDTTIGNIIAEAMAKVGGVITVERA 180
DB 121 DKAVVAVTKELSDITKPTDRDQKIAQVGTISANSDDTTIGNIIAEAMAKVGGVITVERA 180
QY 122 DAAVEEVGNLFPKAKTIQTSAAIAQVGTISANGAAEIGKMTADAMEKVGNEGVITVERA 180
DB 122 DAAVEEVGNLFPKAKTIQTSAAIAQVGTISANGAAEIGKMTADAMEKVGNEGVITVERA 180
QY 181 KGLETTLDVVEGKFDGKGLSPYFVTNPEKMWCELDNPYILCNEKKITSMKMDLPFILEQV 240
DB 181 KGLETTLDVVEGKFDGKGLSPYFVTNPEKMWCELDNPYILCNEKKITSMKMDLPFILEQV 240
QY 241 AKVNRPLIIIAEDVEGEALATLVNKLRGALVAVKAPGFGERRKAMLEIDAILTGGEA 300
DB 241 AKVNRPLIIIAEDVEGEALATLVNKLRGALVAVKAPGFGERRKAMLEIDAILTGGEA 300
QY 301 IFEDRIKLENSVLSLGTAKRVVDKENTIVDQKSEDIKARVQIRAIQAEETSSDY 360
DB 301 IFEDRIKLENSVLSLGTAKRVVDKENTIVDQKSEDIKARVQIRAIQAEETSSDY 360
QY 361 DREKLERLAKVGGVAVIHVGAATEMEKKDVEDALNATRAAEEGIVPGGGTAFV 420
DB 361 DREKLERLAKVGGVAVIHVGAATEMEKKDVEDALNATRAAEEGIVPGGGTAFV 420
QY 421 RSITKVLDDIKPADDDDELALNIIIRRSLEPLRQIAANAGVEGSIIVKVRPKDGFEN 479
DB 421 RSITKVLDDIKPADDDDELALNIIIRRSLEPLRQIAANAGVEGSIIVKVRPKDGFEN 479
QY 480 AASGEYEDLITKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPEPKD---PMPGG 536
DB 480 AASGEYEDLITKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPEPKD---PMPGG 536
QY 537 GGMGGMGMD 545
DB 538 GGMGGMGMD 546
RESULT 12
CH64_RHIME STANDARD; PRT; 545 AA.
ID CH64_RHIME STANDARD; PRT; 545 AA.
AC Q32034;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin 4 (Protein Cpn60 4) (groEL protein 4).
GN GROEL4 OR GROEL4 OR RA0395 OR SMOA0744.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;

```

MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjan M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSYMA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
-----
EMBL; AE007230; AAC65053.1; -.
DR PIR; C95311; C95311.
DR HAMAP; MF 00600; -. 1.
DR InterPro; IPR001844; Chaperlin Cpn60.
DR Pfam; PRF002423; Cpn60/tCF-1.
DR Dfam; PF00118; cpn60_TCPI; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS CPN60; 1.
KW Chaperone; ATP-binding; Heat shock; Multigene family; Plasmid;
KW Complete proteome.
SQ SEQUENCE 545 AA; 57701 MW; 49BA61P23CE70431 CRC64;

Query Match          68.8%; Score 1864; DB 1; Length 545;
Best Local Similarity 70.5%; Pred. No. 4.2e-82;
Matches 387; Conservative 59; Mismatches 97; Indels 6; Gaps 4;

QY      1 MASKETLFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSAKEI 60
Db      1 MAAKEVKFGRSAREKMLRGVDILADAVKVTLGPKGRNVVIDSKSGAPRITKDGVTVAKEI 60

QY     61 ELEDFKENNGAQMVEKPAKPTSDIAGDGGTTATVLAAQAIYREGVKLVAAGRNPMAIKRI 120
Db     61 ELEDFKENNGAQMVRVASKNIDIAIGDGGTTATVLAAQAIYREGAKAVAAGMNPMDIKRI 120

QY    121 DKAAVAVTKELSDITKTPTDQKEIAQVGTSANSDDTTIGNIIAEAMAKVKGKGVITVEEA 180
Db    121 DLAAEAWDKLLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEMOKVGNEGVITVEEA 180

QY    181 KGLETTLDVVEGMKFDRGYLSPFYFTNPCKMVCELDPNVILONEKKITSMKDMLPILEQV 240
Db    181 KTAETELEVVEGQMQRGYLSPFYFTNPCKMWADLEDAPILLHEKLSNLQAAMLPLVLAIV 240

QY    241 AKVNPLLIIVDVEGEALATLVNNLRGALQOVAVKAPGFERRKAMLEDITAITTGGA 300
Db    241 VQTGKPLLIIADVEGEALATLVNNLRGLGLIAVKAAPGFGRDRKAMLEDITAITTGTV 300

QY    301 IFEDRGIKLENVSLSGTAGRVKRWIDKENTTIIVDGAGKSEDIKARVKQIRAQIETTSSDY 360
Db    301 ISEDLGIKLESVTLDMIGRAKKVSITKENTTIIVDGAGQKSDIEGRVAQIKAQIETTSDY 360

QY    361 DREKIQERLAKLVGGVAVTHVGAATETEMKEKORVEDNALNATRAAVEGIVPGGGTAFV 420
Db    361 DREKIQERLAKLAGGVAVTRVGCAETEVEVKEKKDRIDDALNATRAAQEGIVPGG3VALL 420

QY    421 R-SIKVLDDIKPADDDDELAGLINIIRSLPEEPQLQIAANAAGYEGSIIVEKREPDK-DGRGF 478
Db    421 RSSVKI--TVKGGNDQQDAGNVIIRRALQSAPARQIVENAGDRASTIVGKIIFKNTDDFGY 478

```

QY	479	NAASGEYEDLKAGVIDPKKVTTRIALQNAAVASVLLLTTECAIARBPKEPKDMPGGGM	538
Dd	479	NAQTGEYGDMLAMGIIDPVKKVVRTALQDAASVASLITTEAMIAEL--PKDAOAMPGGM	538
QY	539	GGMGGMDGM 547	
Dd	537	GGMGGMDMM 545	
RESULT 13			
ID	CH60_XYLFA	STANDARD;	PRT; 547 AA.
AC	Q9FPF2,		
DT	28-FEB-2003	(Rel. 41, Created)	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DE	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	60 kDa chaperonin (Protein Cpn60) (groEL protein).		
GN	GROL OR GROEL OR XF0615.		
OS	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;		
OC	Xanthomonadaceae; Xylella.		
NCBI_TaxID=2371;			
[1]			
RN	SEQUENCE FROM N.A.		
RP			
RC	STRAIN=945C;		
RX	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,		
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,		
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,		
RA	Faciniani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.B., Kitajima J.P.,		
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moore D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,		
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,		
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,		
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,		
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshuko M.H.,		
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,		
RA	Zago M.A., Zatz M., Meidanis J., Setubal J.C.;		
RT	"The genome sequence of the plant pathogen Xylella fastidiosa."		
RL	Nature 406:151-159(2000).		
CC	-!- FUNCTION: Prevents misfolding and promotes the refolding and		
CC	proper assembly of unfolded polypeptides generated under stress		
CC	conditions (By similarity).		
CC	-!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of		
CC	7 subunits (By similarity).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).		
CC	-!- SIMILARITY: Belongs to the chaperonin (HSP60) family.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC			
EMBL; AF2003907; AAF83425.1; --			
DR	PIR; F82783; F82783		

```
DR HSSP; P06139; 1GRL.
DR HAMAP; MF_00600; ; 1.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 547 AA; 57757 MW; 30BE7F937CA7A9D8 CRC64;

Query Match      68.8%; Score 1862.5; DB 1; Length 547;
Best Local Similarity 68.8%; Pred. No. 4.9e-82;
Matches 366; Conservative 80; Mismatches 97; Indels 5; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVTGLPGKGRNVVIEKSPGSPVITKDGVSVAKEI 60
DB 1 MAKEIIFSEKARSRWVHGNNLANAVKATLPGKGRHVLDKSPGSPVITKDGVSVAKEI 60
QY 61 ELEDKFNNGAQMVEKAPKTSIDAGDGTATVLAQALYREGVKLVAAGRNPMALKRGI 120
DB 61 ELADKFNNGAQMVEKAPKTSIDAGDGTATVLAQALYREGVKLVAAGRNPMALKRGI 120
QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTIGNIIAEAMAKVKGKGVITVEEA 180
DB 121 DKAVIAVTELEKKSIDKPTSDKALIAQVATISANSDESIGNIIAEAMAKVKGKGVITVEEA 180
QY 181 KGLTTLDDVVEGKMFDRGYLSPYFVTNPEKMWCELDNPNYILCNEKKITSMKMDLPILQV 240
DB 181 TTELENDVVEGQMDFRGYSPPYFVNNQSQIVELDNPYILLDFKXISSVRDLTVLDAV 240
QY 241 AKVNRPLLIIAEDVEGEALATLVNKLQALQVAVKAPGFGERRKAMLEDAIILTGGEA 300
DB 241 AKESKPLLIIVAEVEGEALATLVNNGIRGLIKVCAVKAAPGFGDRRKAMLEDAVILTGTV 300
QY 301 IFEDRGIKLENVSLISLTAKRVVIDKENTTVVDGAGKSEDIKARVKQIRAQIETSSDY 360
DB 301 ISEVGLSELEKATSHLGRKVKVSKENTTVIIDGIDNDAINGRVKQIKTQIETSSDY 360
QY 361 DREKLERAKLVGGVAVIHVGAATETEMKEKORVEDALNATRAAVEGIVPGGGTAFV 420
DB 361 DREKLERAKLVGGVAVIKVGAATEVEMKEKARVDDALLATRAAVEGIVPGGGVALI 420
QY 421 RSKVLDDIKPADDDDELAGNIIIRSLRQIAANAGVEGSIIVVEKVRPKDGFGENA 480
DB 421 RAITAISNLKGANEDQTHGQIALRAMEAPLRIRIVANAGEEPIVILNKVKGKDNFGYKA 480
QY 481 ASGEYEDLIKAGVIDPKKVTIRIALQNAASVASILLTTTECAIAEKPKPKMPMP---GGG 537
DB 481 ATGEFGDMVNLGILDPTKVTIRSLQNAASLAGLMIITEAMVAE--APKXDEPTPPAAGG 538
QY 538 MGGMGGM 545
DB 539 MGGMGGM 546

RESULT 14
CH60 RHOMR STANDARD; PRT; 540 AA.
ID CH60 RHOMR
AC QSKC9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL.
OS Rhodothermus marinus (Rhodothermus obamensis).
OC Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC Crenotrichaceae; Rhodothermus.
OX NCBI_Taxid=29549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ITI 376;
RA Thoralfsdottir E.T.T., Backman V.M., Blondal T.,
```

RESULT 15
 CH60 CHRVI STANDARD; PRT; 546 AA.
 AC P31293;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
 GN GROEL OR GROEL OR MOPA.
 OS Chromatium vinosum.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
 OC Chromatiaceae; Allochromatium.
 OX NCBI_TaxID=1049;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=93186721; PubMed=8444812;
 RA Ferreyra R., Soncini F., Viale A.M.;
 RT "Cloning, characterization, and functional expression in Escherichia
 RT coli of chaperonin (groEL) genes from the phototrophic sulfur
 RT bacterium Chromatium vinosum.";
 RL J. Bacteriol. 175:1514-1523(1993).
 RN [2]
 RN CHARACTERIZATION.
 RX MEDLINE=99008942; PubMed=9790891;
 RA Dionisi H.M., Viale A.M.;
 RT "Purification and characterization of Chromatium vinosum GroEL and
 RT GroES proteins overexpressed in Escherichia coli cells lacking the
 RT endogenous groEL operon.";
 RL Protein Expr. Purif. 14:275-282(1998).
 CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
 CC proper assembly of unfolded polypeptides generated under stress
 CC conditions.
 CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 CC 7 subunits.
 CC -1- SURCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M92443; AAA23319.1; -;
 CC PIR; B47073; B47073.
 CC HSP; P06139; IGRL.
 CC HAMAP; MF 00600; -; 1.
 CC InterPro; IPR001844; Chaperin Cpn60.
 CC InterPro; IPR002423; Cpn60/TCF-1.
 CC Pfam; PF00118; cpn60_TCF1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCPI.
 CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 CC KW Chaperone; ATP-binding.
 CC SEQUENCE 546 AA; 57541 MW; 3250141881C04DD6 CRC64;
 DR
 DR PIR; B47073; B47073.
 DR HSP; P06139; IGRL.
 DR HAMAP; MF 00600; -; 1.
 DR InterPro; IPR001844; Chaperin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCF-1.
 DR Pfam; PF00118; cpn60_TCF1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCPI.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 DR KW Chaperone; ATP-binding.
 DR SEQUENCE 546 AA; 57541 MW; 3250141881C04DD6 CRC64;
 SQ

Query Match
 Best Local Similarity 68.4%; Score 1852.5; DB 1; Length 546;
 Matches 374; Conservative 70; Mismatches 101; Indels 3; Gaps 2;

QY 1 MASKEILFDKAREKLSRGYDKLANAVKVTGLGPKGRNVLEKSPGSPVITKGVSAKEI 60
 DB 1 MSKADVKFGGDARVRMMEGVNLANAVKVTGLGPKGRNVLEKSPGSPVITKGVSAKEI 60
 QY 61 ELEDKFNMGAMQVKEVAPKTSIADGTTTATVLAQIYREGVKLVAGRNPMIAIKRGI 120
 DB 61 ELKDKFNMGAMQVKEVAKTSIDAGTTTATVLAQAVRREGKLVAGRNPMIAIKRGM 120
 QY 121 DKAVAVTKELSDITKPTRDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGKGVITVEEA 180
 DB 121 DKAVEAATEELKLSKPCRPMAIAQVGTISANSDDSGTIIAEAMAKVGKGVITVEEDG 180

Db 121 DKAVEAATEELKLSKPCRPMAIAQVGTISANSDDSGTIIAEAMAKVGKGVITVEEDG 180
 QY 181 KGLETTLDVVEGMPDRGYLSFYFTNPENKVMVCELDNPFYILCNEKKITSMKMDLPILQEV 240
 Db 181 TSLQNELDVVEGMPDRGYLSFYFTNPENKVMVCELDNPFYILCNEKKITSMKMDLPILQEV 240
 QY 241 AKVNRPLLIIEADVEGEALATLVNKLKRGALQVAVKAGRGGERKAMLEDTAILTGGEA 300
 Db 241 AKAGKPLLIIEADVEGEALATLVNKLKRGALQVAVKAGRGGERKAMLEDTAILTGGEA 300
 QY 301 IFEDRGIKLENVSLSSLGCTAKRVVIDKENTTVIDGAGKSEDIKARVKOIRAOIBETSSDY 360
 Db 301 ISEEVGLSLEKATLTDLTAKRVQVQKDETTIIDGSGSEIDLKARCEQIRAGVETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGAATEIEMKEKKORVEDALNATRAAEEGIVPGGCTAFV 420
 Db 361 DREKLOERLAKLVGGVAVIHVGAATEIEMKEKKORVEDALNATRAAEEGIVPGGCTAFV 420
 QY 421 RSIKVLDDIKPADDDDELALNIIIRSLPEEPLRQIAANAGYEGSIVVEKREPKGDFGNA 480
 Db 421 RATAAVKDLKGANHDQDVGIARAMEEPLRQIVANAGEEFSVILHKVABGTGNFGYNA 480
 QY 481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLITTECAIAEKPEPKDMP-MEGGGMG 539
 Db 481 ANGEYGDVMEGILDPKTVRSALQNSCVAGLMTTEAMIAD--EPKDDAPAMPGGMG 538
 QY 540 GMGGMGDM 547
 Db 539 DMGGMGMM 546

Search completed: January 28, 2004, 13:16:24
 Job time : 19 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:11:20 ; Search time 39 Seconds
(without alignments)
3625.970 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDKAREKLSRGV.....KDMPPGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phase:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteria:.*
17: sp_archaea:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2708	100.0	548	2	O87888
2	1880	69.4	546	2	O8GBD2
3	1823	67.3	552	2	O8KJ46
4	1819	67.2	546	16	O8FX87
5	1804.5	66.6	545	2	O8GB95
6	1796.5	66.3	549	2	O9AJB5
7	1780.5	65.7	545	16	O8CX48
8	1780.5	65.7	546	2	O8GBB4
9	1768	65.3	539	2	O9RC20
10	1768	65.3	539	2	O31198
11	1754	64.8	539	2	O9EZV4
12	1742	64.3	545	2	O93FU8
13	1736	64.1	538	2	O33688
14	1722.5	63.6	546	2	O8RNU2
15	1714.5	63.3	548	2	O8KIW5
16	1711	63.2	543	2	O8KZN2

17	1710.5	63.2	548	2	O8KIW7
18	1706	63.0	555	2	O9X603
19	1694	62.6	542	16	O8CWW6
20	1691.5	62.5	536	2	O8KIW2
21	1691	62.4	537	2	O8KJ14
22	1687	62.3	545	16	O8CX13
23	1686	62.3	540	16	O8CX22
24	1685	62.2	531	2	O8GBD0
25	1683.5	62.2	540	16	O8CX00
26	1681	62.1	536	2	O8KIW4
27	1679.5	62.0	537	2	O8KIW3
28	1679.5	62.0	537	2	O8KIW3
29	1676.5	61.9	537	2	O8GBB6
30	1675	61.9	534	2	O8GBB6
31	1672.5	61.8	548	2	O08499
32	1671.5	61.7	548	2	O08500
33	1670	61.7	531	2	O8GBC6
34	1669	61.6	531	2	O8GBC4
35	1669	61.6	531	2	O8GBC0
36	1668	61.6	534	2	O8KIY3
37	1667	61.6	534	2	O8KJ12
38	1663.5	61.4	540	2	O9X4R5
39	1663	61.4	531	2	O8GBC2
40	1663	61.4	535	2	O8KJ08
41	1660	61.3	503	2	O8J271
42	1660	61.3	532	2	O8GBC8
43	1659	61.3	532	2	O8GBB8
44	1648.5	60.9	539	2	O8GJ00
45	1644	60.7	532	16	O8CWW0

ALIGNMENTS

RESULT 1

O87888 PRELIMINARY; PRT; 548 AA.
AC O87888, 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE GroEL/HSP60 homolog (60 kDa chaperonin) (Protein Cpn60) (groEL protein).
OS Lawsonia intracellularis.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfotribionales; Desulfotribionaceae; Lawsonia.
CX NCBI_TaxID=29546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98386497; PubMed=9720028;
RA Dale C.J.H., Moses E.K., Ong C.C., Morrow C.J., Reed M.B., Hasse D., Strugnell R.A.;
RT Identification and sequencing of the groE operon and flanking genes of Lawsonia intracellularis: use in phylogeny.
RL Microbiology 144:0-0(10).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL; U45241; AAC36500.1; -.
DR HSP; P06139; IGR1.
DR InterPro; IPR001844; Chaperin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN CPN60; 1.
KW ATP-binding; Chaperone; Plasmid.
SQ SEQUENCE 548 AA; 58605 MW; 6380C431E663E498 CRC64;

O8kiw7 buchnera sp
O9x603 primary end
O8CWW6 streptococ
O8kiw2 buchnera sp
O8kjl4 streptococ
O8cx13 oceanobacil
O8gbd0 streptococ
O8cx00 enterococ
O8kiw4 buchnera sp
O8kiw3 streptococ
O8kiw3 buchnera sp
O8gbb6 streptococ
O08499 rhopalosiph
O08500 sitobion av
O8gbc6 enterococ
O8gbc4 enterococ
O8gbc0 enterococ
O8kiy3 streptococ
O8kjl2 streptococ
O9x4r5 streptococ
O8gbc2 enterococ
O8kj08 streptococ
O87271 bartonella
O8gbc8 enterococ
O8gbb8 enterococ
O8gju0 fusobacteri
O8cww0 vibrio vuln

Query Match 100.0%; Score 2708; DB 2; Length 548;
 Best Local Similarity 100.0%; Pred. No. 1.3e-124;
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVLGPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 DB 1 MASKEILFDKAREKLSRGVDKLANAVKVLGPKGRNVVIEKSPGSPVITKDGVSVAKEI 60
 QY 61 ELEDKFNMGQAQVKEVAPKTSIDAGDGTITATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 DB 61 ELEDKFNMGQAQVKEVAPKTSIDAGDGTITATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 QY 121 DKAVAVVTKELSDITKPTRDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGGVITVEEA 180
 DB 121 DKAVAVVTKELSDITKPTRDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGGVITVEEA 180
 QY 181 KGLETTLDVVEGKMPDRGVLSPYFVTNPEKMWCELDNPIYILCNEKKITSMKMDLPILQV 240
 DB 181 KGLETTLDVVEGKMPDRGVLSPYFVTNPEKMWCELDNPIYILCNEKKITSMKMDLPILQV 240
 QY 241 AKVNRPLIIIAEDVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 DB 241 AKVNRPLIIIAEDVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 QY 301 IPEDRGIKLENVSLSGTAKRVVIDKENTTIIVDAGKSEDIKARVKQIRAOIETSSDY 360
 DB 301 IPEDRGIKLENVSLSGTAKRVVIDKENTTIIVDAGKSEDIKARVKQIRAOIETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVERGIVPGGTAFF 420
 DB 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVERGIVPGGTAFF 420
 QY 421 RSIKVLDDIKPADDDDELIGINIIRRSLEPRLQIAANAGYEGSVIVVEKVRPKDGFGFNA 480
 DB 421 RSIKVLDDIKPADDDDELIGINIIRRSLEPRLQIAANAGYEGSVIVVEKVRPKDGFGFNA 480
 QY 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLITTECAIAEKPKKMDMPGGMGG 540
 DB 481 ASGEYEDLIKAGVIDPKVTRIALQNAASVASLLITTECAIAEKPKKMDMPGGMGG 540
 QY 541 MCGMDGMY 548
 DB 541 MCGMDGMY 548

RESULT 2
 Q8GBD2 PRELIMINARY; PRT; 546 AA.
 AC Q8GBD2; (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE groEL protein.
 GN GROEL.
 OS Acetobacter aceti.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Acetobacteraceae; Acetobacter.
 OX NCHI_TaxID=435;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=IFO3283;
 RA Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,
 RA Yanagida F.;
 RT "Cloning and Characterization of groEL Operon in Acetobacter aceti."
 RL J. Biosci. Bioeng. 94:140-147(2002).
 DR EMBL; AB081586; BAC16232.1; --
 SQ SEQUENCE 546 AA; 58106 MW; 3E476AF42F2CD5CC CRC64;

Query Match 69.4%; Score 1880; DB 2; Length 546;
 Best Local Similarity 69.8%; Pred. No. 3.5e-84;
 Matches 381; Conservative 69; Mismatches 94; Indels 2; Gaps 2;

QY 1 MASKEILFDKAREKLSRGVDKLANAVKVLGPKGRNVVIEKSPGSPVITKDGVSVAKEI 60

DB 1 MAADKDFGADARQMRGAGVDILADAVKVLGPKGRNVVILDKSGAPRIITKDGVSVAKEI 60
 QY 61 ELEDKFNMGQAQVKEVAPKTSIDAGDGTITATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 DB 61 ELADKFNMGQAQVKEVASKTNDIAGDGTITATVLAQAIYREGCHKAAGNPNMDLKRGI 120
 QY 121 DKAVAVVTKELSDITKPTRDQKEIAQVGTISANSDDTTIGNIIAEAMAKVGGVITVEEA 180
 DB 121 DKAVAVVIEELKNAKVTTPAETAQVGTISANGESEIGQMISEAMQKVGSEGVITVEEA 180
 QY 181 KGLETTLDVVEGKMPDRGVLSPYFVTNPEKMWCELDNPIYILCNEKKITSMKMDLPILQV 240
 DB 181 KHFQTELDVVEGKMPDRGYISPYFVTNPEKMTADLENPIYILHEKLSLOPMLLESV 240
 QY 241 AKVNRPLIIIAEDVEGEALATLVNKLKRGALQVAVKAPGFGERRKAMLEDAITLTGGEA 300
 DB 241 VQSGRPLIIIAEDVDGEALATLVNKLKRGGLKIAAVKAPGFGDRNRNMLEDAITLTGGQV 300
 QY 301 IPEDRGIKLENVSLSGTAKRVVIDKENTTIIVDAGKSEDIKARVKQIRAOIETSSDY 360
 DB 301 ISEDLGKLETVTLNMLGTAKKVHIDKENTTIIVDAGKADDIKGRVKQIRAOIETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVERGIVPGGTAFF 420
 DB 361 DREKLOERLAKLVGGVAVIRVGGSTEVEVEKRDVDDALHATRAAVERGIVPGGTTALA 420
 QY 421 RSIKVLDDIKPADDDDELIGINIIRRSLEPRLQIAANAGYEGSVIVVEKVRPKD-GFGFN 479
 DB 421 RATLKLEGLHYHNDQVGGDIIRRALQAPLQIAHNAAGEDGAVIANKYLENSDYNFGFD 480
 QY 480 AASGEYEDLIKAGVIDPKVTRIALQNAASVASLLITTECAIAEKPEPKKMDMPGGMGG 539
 DB 481 AQAGEYKLVVAGIIPAKVWRTALQDAASVASGLITTEAMVAERPE-KKAPAGPDDMG 539
 QY 540 GMGMD 545
 DB 540 GMGMD 545

RESULT 3

Q8KJ46 PRELIMINARY; PRT; 552 AA.
 ID Q8KJ46;
 AC Q8KJ46; (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 23, Last annotation update)
 DE Probable chaperonin groEL DP protein (60 kDa chaperonin) (Protein
 DE Cpn60) (groEL protein).
 GN GROEL.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCHI_TaxID=381;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=R7A;
 RC MEDLINE=21999272; PubMed=12003951;
 RA Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J.,
 RA Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,
 RA Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;
 RT "Comparative sequence analysis of the symbiosis island of
 Mesorhizobium loti strain R7A."
 RL J. Bacteriol. 184:3086-3095(2002).
 CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 DR EMBL; AL672115; CAD31231.1; --
 DR InterPro; IPR001844; Chaperin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.

DR Pfam; PF00118; cpn60_TCP1; 1.
 DR PRINTS; PRO0298; CHAPERONIN60.
 DR PRINTS; PRO0304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONINS_CP60; 1.
 KW ATP-binding; Chaperone
 SQ SEQUENCE 552 AA; 57747 MW; 7D0A8C53CD7FA6AB CRC64;

Query Match 67.3%; Score 1823; DB 2; Length 552;
 Best Local Similarity 67.8%; Pred. No. 2.2e-81;
 Matches 374; Conservative 73; Mismatches 97; Indels 8; Gaps 4;

QY 1 MASKEILLFDKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSFSGSPVITKDGVSVAKEI 60
 DB 1 MAADKDVKFSRDAREKMLRGVILADAVKVTLPKGRNVVIEKSFSGSPVITKDGVSVAKEI 60
 QY 61 ELEKFFENMGQAQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 DB 61 ELEKFFENMGQAQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGKGVITVEEA 180
 DB 121 DLAVTDVVATLTKNAKKIKTSEEVAVQVGTISANGDESVGKMAEAMQKVGNEGVITVEEA 180
 QY 181 KGLTLLDVVEGKMFDRGYLSPYFVTNPEKMWCELDNPNYILCNBKKTITSMKDMPLILEQV 240
 DB 181 KTAETELEVVEGQFDRGYLSPYFVTNPEKMWCELDNPNYILCNBKKTITSMKDMPLILEQV 240
 QY 241 AKVNRPLIITAEVDVEGALATLVNKLKRGALQVAVKAPGGERKAMLEIDIALITGGEA 300
 DB 241 VQTSKPLIITAEVDVEGALATLVNKLKRGALQVAVKAPGGERKAMLEIDIALITGGEA 300
 QY 301 IFEDRGKLENVSLSSITGAKRVVIDKENTTIVDAGKSEDIKARVKQIRAOIEETSSDY 360
 DB 301 ISEDLGKLENVSLSSITGAKRVVIDKENTTIVDAGKSEDIKARVKQIRAOIEETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGATETEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
 DB 361 DREKLOERLAKLVGGVAVIHVGATETEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
 QY 421 RSTKVLDDIKPADDELAGNIIRSLLEPIRLQIAANAGVEGSIIVVEKVPKDG-FGPN 479
 DB 421 RASLSINAV-GANSDDQAGNIVRRALQAPARQIAANAGAEASIVAGKILENKATGEGYN 479
 QY 480 AASGEYEDLIKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPEPKK---DMP--M 533
 DB 480 AQTGEYEDLIKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPEPKK---DMP--M 533
 QY 534 PGCGMGGMGMD 545
 DB 540 GGGGGMGGMGMD 551

RESULT 4
 Q8FX87 PRELIMINARY; PRT; 546 AA.
 AC Q8FX87;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 GN GROEL OR BRA0195.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,

RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.,
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 DR EMBL; AE014521; AAN33401.1; -
 DR TIGR; BRA0195; -
 KW Complete proteome.
 SQ SEQUENCE 546 AA; 57501 MW; 2E4046EDC121A30A CRC64;

Query Match 67.2%; Score 1819; DB 16; Length 546;
 Best Local Similarity 68.7%; Pred. No. 3.4e-81;
 Matches 377; Conservative 64; Mismatches 100; Indels 8; Gaps 5;

QY 1 MASKEILLFDKAREKLSRGVDKLANAVKVTLPKGRNVVIEKSFSGSPVITKDGVSVAKEI 60
 DB 1 MAADKDVKFSRDAREKMLRGVILADAVKVTLPKGRNVVIEKSFSGSPVITKDGVSVAKEI 60
 QY 61 ELEKFFENMGQAQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 DB 61 ELEKFFENMGQAQVKEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGNPNMAIKRGI 120
 QY 121 DKAVAVTKELSDITKPTDQKEIAQVGTISANSDDTTIGNIIAEAMAKVKGKGVITVEEA 180
 DB 121 DLAVNEVVAELLKKAKKINTSEEVAVQVGTISANGAEIGKMAEAMQKVGNEGVITVEEA 180
 QY 181 KGLTLLDVVEGKMFDRGYLSPYFVTNPEKMWCELDNPNYILCNBKKTITSMKDMPLILEQV 240
 DB 181 KTAETELEVVEGQFDRGYLSPYFVTNPEKMWCELDNPNYILCNBKKTITSMKDMPLILEQV 240
 QY 241 AKVNRPLIITAEVDVEGALATLVNKLKRGALQVAVKAPGGERKAMLEIDIALITGGEA 300
 DB 241 VQTSKPLIITAEVDVEGALATLVNKLKRGALQVAVKAPGGERKAMLEIDIALITGGEA 300
 QY 301 IFEDRGKLENVSLSSITGAKRVVIDKENTTIVDAGKSEDIKARVKQIRAOIEETSSDY 360
 DB 301 ISEDLGKLESVTLDMGRKKVSIKENTTIVDAGKAEIDARVQIQKQIEETSSDY 360
 QY 361 DREKLOERLAKLVGGVAVIHVGATETEMKEKDRVEDALNATRAAEEGIVPGGGTAFV 420
 DB 361 DREKLOERLAKLVGGVAVIRVGATEVEVEKDRVDALNATRAAEEGIVAGGGTALL 420
 QY 421 R-SIKVLDDIKPADDELAGNIIRSLLEPIRLQIAANAGVEGSIIVVEKVPKDG-FGFGF 478
 DB 421 RASTKI--TAGVNAQOEAGINIVRRALQAPARQITTNAGEEASIVGKILENTSETFGY 478
 QY 479 NAASGEYEDLIKAGVIDPKKVTIRIALQNAASVASILLITTECAIAEKPEPKKMPMPG--G 536
 DB 479 NTANGEYGDILISLIGVDPKVVRTALQNAASVAGLLITTEAMIAEL--PKKDAAPAGMPG 536
 QY 537 GMGGMGMD 545
 DB 537 GMGGMGMD 545

RESULT 5
 Q8GB95 PRELIMINARY; PRT; 545 AA.
 AC Q8GB95;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 GN 60 kDa chaperonin groEL (fragment).
 OS Helicobacterium mobilis.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
 OC Helicobacter.
 OX NCBI_TaxID=28064;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22337798; PubMed=12446909;
 RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
 RA Blankenship R.E.,
 RT "Whole-genome analysis of photosynthetic prokaryotes."

Science 298:1616-1620 (2002).
[2]
SEQUENCE FROM N.A.
Liolios K.G., Chu L., Ostrovskaya O., Mendiyaeva N., Koukharenko V.,
Gerdes S., Kyripides N., Overbeek R.,
Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
EMBL; AY142910; AAN87514.1; -;
NON TER 545
SEQUENCE 545 AA; 57927 MW; B9F6DB02DBC18383 CRC64;
Query Match 66.6%; Score 1804.5; DB 2; Length 545;
Best Local Similarity 65.9%; Pred. No. 1.7e-80;
Matches 360; Conservative 84; Mismatches 99; Indels 3; Gaps 3;
3 SKEILFDKAREKLSRGVDKLANAVKVTLPGRNVVIEKSPGSPVITKDGVSVAKEI 62
2 AKMTVFEEARRALEKGVNTLAEAVRVTLGPKGRNVVLEKFKGSPVITNDGVTIAKEI 61
63 EDKFENMGQAWKVEAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAIKRGIDK 122
62 ENPIENMGQALVKEVATKNDVAGDGTATVLAQAIYREGKVNVAAGNPMVIRGIEK 121
123 AVAVTKELSDITKTRDQKEIAQVGTISANSDDTIGNIIABAMAKVKGKGVITVEAKG 182
122 AVATCVSEIKNLKSPVESKEIAQVASISA-ADATIGSLIAEAMEKVGKGVITVEESKG 180
183 IETTLVVEGKFDGKYLSPYFVTNPKEKMCVCELDNPYILCNEKKITSMKMLPILQVAK 242
181 FSTDLEVEGKMFDRGYISPYMITDPDKMEAVLNDPYILITDKISAIKDLFVLIRIVQ 240
243 VNRPLLIITAEVGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEIDAILTGGEATF 302
241 SGKQLMIISDIEGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEIDAILTGGRVVS 300
303 EORGKLENVSLSSLTAKRVVDKENTTIVDAGKSEDIKARVKQIRAQIEETSSDYDR 362
301 EGVGLKLENATMDLGRQVRSKEETIIVDAGSQDDIKARIAIRQYEEESTSEFDK 360
363 EKIQLERLAKLVGGVAVIHVGAATETEMKKKRDVEDALNATRAAEEGIVPGGTAFTVS 422
361 EKLQERLAKLVGGVAVIQGAAVETELKEKRLIEDALNATRAAEEGIVPGGTAFTVS 420
423 IKVLDDIKPADDDDELALNIIIRSLPELRQIAANAGYEGSIIVVEKVRPKDGFNAAS 482
421 OKALDEVEVPAGDEATGVALLKESLEPLRQIANNAGYEGSVVVEKVKMLPKQGGFNAAT 480
483 GEYEDLIKAGVIDPKKTRIALQNAASVASLLITTECAIAEKPEPKKMPGPGMG 542
481 EYVEDMIAAGIVDPKTRIALQNAASIASIAAMLITTEAIVADKPE-KKEAPGMPGPGMG 539
543 GMD-GM 547
540 GMDMG 545
RESULT 6
Q9AJB5 PRELIMINARY; PRT; 549 AA.
AC Q9AJB5; 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE groEL protein (60 kDa chaperonin) (protein Cpn60).
GN GROEL.
OS Pseudomonas sp. PslM3.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Pseudoalteromonas.
OX NCBI_TaxID=87791;
RN [1]
RN SEQUENCE FROM N.A.
RN STRAIN=PSIM3;
RA Kurusu Y., Nakamura T.;
RT "Molecular chaperone of Psychrophile.";

Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CONDITIONS (BY SIMILARITY).
-!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
7 SUBUNITS (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
EMBL; AB057417; BAB39465.1; -;
HSSP; P06139; 1GR1.
DR InterPro; IPR001844; Chaprinin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 549 AA; 57346 MW; 7AFCC2E8A2F736AF CRC64;
Query Match 66.3%; Score 1796.5; DB 2; Length 549;
Best Local Similarity 65.7%; Pred. No. 4.3e-80;
Matches 360; Conservative 79; Mismatches 108; Indels 1; Gaps 1;
QY 1 MASKELFDKAREKLSRGVDKLANAVKVTLPGRNVVIEKSPGSPVITKDGVSVAKEI 60
DB 1 MAKEVLFAGDARAKMLTGVNLLANAVKVTLPGRNVVLDKSPGSPVITKDGVSVAKEI 60
QY 61 ELEDKFENMGQAWKVEAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAIKRG 120
DB 61 ELEDKFENMGQAWKVEAPKTSIDAGDGTATVLAQAIYREGVKLVAAAGRNPMIAIKRG 120
QY 121 DRAVAVTKELSDITKTRDQKEIAQVGTISANSDDTIGNIIABAMAKVKGKGVITVEE 179
DB 121 DRAVAVTKELSDITKTRDQKEIAQVGTISANSDDTIGNIIABAMAKVKGKGVITVEE 180
QY 180 AKGLTTLVVEGKFDGKYLSPYFVTNPKEKMCVCELDNPYILCNEKKITSMKMLPILQ 239
DB 181 GOSLENELDVVEGKFDGKYLSPYFVTNPKEKMCVCELDNPYILCNEKKITSMKMLPILQ 240
QY 240 VAKVNRPLLIITAEVGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEIDAILTGGE 299
DB 241 VAKASKPLLIITAEVGEALATLVNKLKRGALQVAVKAPGFGRRKAMLEIDAILTGGE 300
QY 300 AIFEDGKLENVSLSSLTAKRVVDKENTTIVDAGKSEDIKARVKQIRAQIEETSSD 359
DB 301 VISEETGLEKATVEDLTAKRVVITKDDTTIIDGAGEEGITGRVQIKAQIEETSD 360
QY 360 YDEKLOERLAKLVGGVAVIHVGAATETEMKKKRDVEDALNATRAAEEGIVPGGTAFT 419
DB 361 YDEKLOERLAKLVGGVAVIKVGAATEIEMKKKRDVEDALNATRAAEEGIVPGGTAFT 420
QY 420 VRSIKVLDDIKPADDDDELALNIIIRSLPELRQIAANAGYEGSIIVVEKVRPKDGFNF 479
DB 421 VRAASKVLELLGDNEDQNHGKVALRAMEAPLQIVTNAGDEASVVVNAVKGEGNYGN 480
QY 480 AASGEYEDLIKAGVIDPKKTRIALQNAASVASLLITTECAIAEKPEPKKMPGPGMG 539
DB 481 AATGEYNDMIEGILDPTKTRIALQNAASVASLLITTECAIAEKPEPKKMPGPGMG 540
540 GMDMG 547
541 GMDMG 548
RESULT 7
Q8CX48 PRELIMINARY; PRT; 545 AA.
ID Q8CX48
AC Q8CX48; 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chaperonin groEL.
GN GROEL OR S00704.
OS Shewanella oneidensis.

Q9RC20;	Db	534 DMGM 538
AC	031198	PRELIMINARY; PRT; 546 AA.
DT	01-JAN-1998	(TReMBLrel. 05, Created)
DT	01-MAY-2000	(TReMBLrel. 13, Last sequence update)
DT	01-OCT-2002	(TReMBLrel. 22, Last annotation update)
DE	Protein Cpn60	(60 kDa chaperonin) (groEL protein).
GN	Cpn60.	
OS	Bacillus sp. MS.	
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
OX	NCBI_TaxID=96470;	
ON	[1]	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=MS;	
RA	Kondo A.; Yoshida M.;	
RT	"Cpn60/10 from Bacillus strain MS.";	
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.	
CC	-!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).	
CC	-!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).	
CC	-!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.	
CC	EMBL; AB028452; BAA8110.1; -.	
DR	HSP; P06139; 1GRL.	
DR	InterPro; IPR001844; Chaperin Cpn60.	
DR	InterPro; IPR002423; Cpn60/TCP-1.	
DR	Pfam; PF00118; cpn60 TCP1; 1.	
DR	PRINTS; PR00298; CHAPERONIN60.	
DR	PRINTS; PR00304; TCOMPLEXTCP1.	
DR	PROSITE; PS00296; CHAPERONINS_CPN60; 1.	
KW	ATP-binding; Chaperone.	
SW	SEQUENCE 539 AA; 57346 MW; 95626C980E30C46F CRC64;	
QY	Query Match	55.3%; Score 1768; DB 2; Length 539;
Db	Best Local Similarity	65.0%; Pred. No. 1e-78;
QY	Matches	354; Conservative 85; Mismatches 98; Indels 8; Gaps 3;
QY	3	SKEILFPAKAREKLSRGVDKLANAVKVTGLGPKGNVVIKSFSGPVITKDGVSVAKEIEL 62
Db	2	AKETIFSEEARAMLRGVDKLANAVKVTGLGPKGNVVIKSFSGPVITKDGVSVAKEIEL 61
QY	63	EDPENNGAQMVEVAPKTSIDAGDGTATVLAQAIYREGVKLVAAGRNPMALRGIDK 122
Db	62	EDPENNGAKLVAEVASKTNDVAGDGTATVLAQAIYREGVKLVAAGRNPMALRGIDK 121
QY	123	AVAVTVELSDITKPTDQKEIAQVGTISANSDDTIGNITAEAMAKVGKGVITVEEAKG 182
Db	122	AVAVAVELKSAISPKIGKESIAQVAALISA-ADBEVGOLIAEAMERVGNDGVITVEESKG 180
QY	183	LETTLDDVVEGKMFDRGYSLPYFTNPEKMWCELDNPNYILCNEKITSKMKMLPILEQVAK 242
Db	181	FTTELDVVEGQFDRGYSVPYITDTEKWEAVLENPNYILITDKLSINQDILPILEQVQV 240
QY	243	VNRPLIIAEDVEGEALATLVNKLRLGALQVAVKAPGFGERRKAMLEDAIITGGGEAIF 302
Db	241	QKGPLIIAEDIEGALATLVNKLRLGTFVAVKAPGFGDRRKMALIEDAIITGGGEVIS 300
QY	303	EDRGIKLENVSLISGTAKRVVVDKENTTIIVDGAQSEDIKARVKQIRAIQTEETSSDYDR 362
Db	301	BELEGELASATIASLGRASKVVTENTTIIVDGAQSEDIKARVKQIRAIQTEETSEFDR 360
QY	363	EKLQERLAKVGGVAVIHVGAATEMETEKEDORVEDALNATRAAEEGIVPGGGTAFVRS 422
Db	361	EKLQERLAKVGGVAVIHVGAATEMETEKEDORVEDALNATRAAEEGIVPGGGTAFVRS 420
QY	423	IKVLDDIKPADDELALNIIIRSLPEELRQIAANAGYEGSVIVBKVREPKDGFNFNAAS 482
Db	421	YNKVAALIE-AEGDEATGVKIVLRAIEEPVRLQIAANAGYEGSVIVBKVREPKDGFNFNAAT 479
QY	483	GEYEDLIKAGVIDPKVTRIALQNASVASLILLITTECAIAEKPEPKMPPGGGGMG 542
Db	480	GEVMDIEAGVIDPKVTRIALQNASVASLILLITTECAIAEKPEPKMPPGGGGMG 542
QY	543	GMDGM 547

Db 301 EELGRELKSTTIASLRASKVVTWTTTIVEGAGSDRIKARINOIRAQLETTTSEFDR 360

QY 363 EKLQERLAKLVGGVAIVHVGAAATETEMKEKKDRVEDALNATAAEEGIVPGCGTAFVRS 422

Db 361 EKLQERLAKLVGGVAIVHVGAAATETELKERRURIEDALDSTRAAAVEEGIVAGGTMNV 420

QY 423 IKVLDDIKPADDELALAGLNIIRRSLEELPQIAANAGYEGSVVEKVRPKDGFNGFAAS 482

Db 421 YSKVAAIE-AGEDEATGVKVLRAEETPVQIQANAGLEGSVIVRELKTEKGIGENAAI 479

QY 483 GEYEDLLIAGVLDPKVTRIALONAAASVASLLLTTECAIAEKPEPKMPGGGGMG 542

Db 480 GEWNMMIAGIVDPKTVRSALONAAASVAAAMPLTTEAVVADKPEENK-----GGNGMP 533

QY 543 GMDGM 547

Db 534 DMGGM 538

RESULT 12

Q93FU8 PRELIMINARY; PRT; 545 AA.

AC Q93FU8

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE groEL-like protein (60 kDa chaperonin) (protein Cpn60) (groEL

DE protein).

GN GROEL

OS Enterobacter aerogenes (Aerobacter aerogenes).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Enterobacter.

ON NCBI_TaxID=548;

FX [1]

RP SEQUENCE FROM N.A.

RA Yoshida N., Oeda K., Watanabe E., Mikami T., Fukita Y., Nishimura K.,

RA Komai K., Matsuda K.;

RT "A GroEL homolog produced by endosymbionts in antlions as a paralytic

RT toxin against insects.";

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND

CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS

CC CONDITIONS (BY SIMILARITY).

CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF

CC 7 SUBUNITS (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

DR EMBL: AF306521; AAL09389.1; -

DR InterPro: IPR001844; Chaprin Cpn60.

DR InterPro: IPR002423; Cpn60/TCF-1.

DR Pfam: PF00118; cpn60 TCP1; 1.

DR PRINTS; PR00298; CHAPERONIN60.

DR PRINTS; PR00304; TCOMPLEXTCP1.

DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.

KW ATP-binding; Chaperone.

SQ SEQUENCE 545 AA; 57023 MW; DB3B7A017DEDA85F CRC64;

Query Match 64.3%; Score 1742; DB 2; Length 545;

Best Local Similarity 64.7%; Pred. No. 1.9e-77;

Matches 353; Conservative 73; Mismatches 116; Indels 4; Gaps 2;

QY 1 MASKEILLFPAKAREKLSRGVDKLANAVKVTGLPGKGRNVVIEKSFSPVITKDGVSVAKEI 60

Db 1 MAADKVFNGDARVVKMLRGVNLADAVKVTGLPGKGRNVLDKSFCAPTITKDGVSVAKEI 60

QY 61 ELEDKFNNGAOMKVEAPKTSIDAGDGGTTATVLAQAIYREGVKLVAAGRNPMIAKRG 120

Db 61 ELEDKFNNGAOMKVEASKANDAAGDGGTTATVLAQAIYREGKLVAAAGMNPMDLKRG 120

QY 121 DKRAVAVTKELSDITKPTRDOKEIAQVGTISANSTDTTIGNITAEAMAKYKCGVITVEA 180

Db 121 DKRAVAAVEELKALSVPCSDSKAIAQVGTISANSDTETVKGLIAEAMDVKYKGVITVEDG 180

QY 181 KGLETTLDVVGKMFDRGYLSFYFTNPEKMWCELDNPNYILCNEKKITSMKMDLPILCOV 240

Db	181	TGLQDELVDVVEGMOFDRGYLSPYFINKPETGAVELESFFILLADKKLSNIREMLPVLEAV	240
Qy	241	AKVNRPLIIIAEDVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAIILTGGEA	300
Db	241	AKAGKPLIIIAEDVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAIILTGGEA	300
Qy	301	IFEDRGKLENVSSISLGTAKRVVIDKENTTIIVDGAAGSEDIKARVKQIRAIETSSDY	360
Db	301	ISEEIGMELEKATLEDLGQAKRVVINKDTTIIIDVGDEAAIQGRVTOIQOIEATSDY	360
Qy	361	DREKLOERLAKLVGVAVIHVGAATEETEMKEDKORVEDALNATRAAAVEEIVPGGGTAFV	420
Db	361	DREKLOERLAKLVGVAVIHVGAATEETEMKEDKORVEDALNATRAAAVEEIVPGGGTAFV	420
Qy	421	RSIKVLDDIKPADDDDELALNIIIRSLPEERLQIAANAGYEGSIIVVEKVRPEKDGFGNA	480
Db	421	RVASKIAGLKQNEQNVGKIVKALAEAPLQIIVLNCGEPSVAVNTVKAGDGNVGYNA	480
Qy	481	ASGEYEDLIKAGVIDPKVTRIALQNAASVASILLTTECAIAEKPEPKDMPG--GGM	538
Db	481	ATEEYGNMIDMGILDPTKVRSAQVAAVAGIITTECMVTDL--PKSDAPDLGAAGGM	538
Qy	539	GGMGGM 544	
Db	539	GGMGGM 544	
RESULT 13			
ID	O33688	PRELIMINARY; PRT; 538 AA.	
AC	O33688;		
DT	01-JAN-1998 (Tremblrel. 05, Created)		
DT	01-JAN-1998 (Tremblrel. 05, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	groEL protein (60 kDa chaperonin) (Protein Qpn60) (Fragment).		
GN	GROEL.		
OS	primary endosymbiont of <i>Sitophilus oryzae</i> .		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae.		
OX	NCBI_TaxID=2342;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SFR;		
RC	MEDLINE=98042468; PubMed=9367844;		
RA	Charles H., Haddi A., Guillaud J., Nardon C., Nardon P.;		
RT	"A molecular aspect of symbiotic interactions between the weevil		
RT	<i>Sitophilus oryzae</i> and its endosymbiotic bacteria: over-expression of a		
RT	chaperonin."		
RL	Biochem. Biophys. Res. Commun. 239:769-774(1997).		
CC	-1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND		
CC	PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS		
CC	CONDITIONS (BY SIMILARITY).		
CC	-1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF		
CC	7 SUBUNITS (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.		
DR	EMBL; AF005236; AAB97670.1; -		
DR	HSP; P06139; 1JON.		
DR	InterPro; IPR001844; Chaperin Cpn60.		
DR	InterPro; IPR002423; Cpn60/TCP-1.		
DR	Pfam; PF00118; cpn60_TCP1; 1.		
DR	PRINTS; PRO0298; CHAPERONIN60.		
DR	PRINTS; PRO0304; TCOMPLEXTCPI.		
DR	PROSITE; PS00296; CHAPERONINS_CPN60; 1.		
KW	ATP-binding; Chaperone.		
FT	NON_TER 538 538.		
SQ	SEQUENCE 538 AA; 56683 MW; EFD17668COF09D15 CRC64;		
Query Match	64.1%; Score 1736; DB 2; Length 538;		
Best Local Similarity	64.0%; Pred. No. 3.7e-77;		
Matches 348; Conservative	81; Mismatches 109; Indels 6; Gaps 2;		
Qy	1	MASKIILFDKAREKLSRGVDKLANAVKTLGPKGRNVVIEKSGFSPVITKDGVSVAKEI	60

Db	1	MAAKDVFPGNDARVKMLRGVNVVLADAVKVLGPKGRNVVLDKSGFAPVITKDGVSVAKEI	60
Qy	61	ELEDKFNMGAMQVKEVAPKTSIDIAAGDGTITATVLAQAIYREGVKLIYAAGRNPMIAIKRGI	120
Db	61	ELEDKFNMGAMQVKEVASKANDAGDGTITATVLAQSIYVNEGLKAVAAGNPMIDLKRGII	120
Qy	121	DKAVAVVTKELSDITKTPROKEIAQVGTISANSDDTTIGNIIAEMAKVKGGVITVEEA	180
Db	121	DKAVIAAEEELKLSVPCSDSKALIAQVGTISANSDETGTILIAEMAKVKGGVITVEEG	180
Qy	181	KGLTTLDDVVEGKFDGRLSPYFVNTPEKVMVCELDNPYILCNEKITSMDMLPILPQV	240
Db	181	SGLQDELVDVVEGMOFDRGYLSPYFVNPETGAI ELESFFILLADKKLSNIREMLPVLEAV	240
Qy	241	AKVNRPLIIIAEDVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAIILTGGEA	300
Db	241	AKAGKPLIIIAEDVEGEALATLVNKLKALQVAVKAPGFGERRKAMLEDAIILTGGEA	300
Qy	301	IFEDRGKLENVSSISLGTAKRVVIDKENTTIIVDGAAGSEDIKARVKQIRAIETSSDY	360
Db	301	ISEEIGMELEKATLEDLGQAKRVVINKDTTIIIDVGDKALIDSRTVQINQORDEATSDY	360
Qy	361	DREKLOERLAKLVGVAVIHVGAATEETEMKEDKORVEDALNATRAAAVEEIVPGGGTAFV	420
Db	361	DREKLOERLAKLVGVAVIHVGAATEETEMKEDKORVEDALNATRAAAVEEIVPGGGTAFV	420
Qy	421	RSIKVLDDIKPADDDDELALNIIIRSLPEERLQIAANAGYEGSIIVVEKVRPEKDGFGNA	480
Db	421	RVANSIAELRGNEQNVGKIVKALAEAPLQIIVLNCGEPSVAVNTVKAGDGNVGYNA	480
Qy	481	ASGEYEDLIKAGVIDPKVTRIALQNAASVASILLTTECAIAEKPEPKDMPG--GGM	540
Db	481	ATEEYGNMIDMGILDPTKVRSAQVAAVAGIITTECMVTDL--OPKEDK-----DLGG	534
Qy	541	GGMGGM 544	
Db	535	GGMGGM 538	
RESULT 14			
Q8RN2			
ID	Q8RN2	PRELIMINARY; PRT; 546 AA.	
AC	Q8RN2;		
DT	01-JUN-2002 (Tremblrel. 21, Created)		
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	Heat shock protein B subunit (60 kDa chaperonin) (Protein Cpn60)		
DE	(groEL protein).		
GN	HSPB.		
OS	Helicobacter pylori.		
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;		
OC	Helicobacteraceae; Helicobacter.		
OX	NCBI_TaxID=210;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CH-CTX1;		
RC	MEDLINE=20123532; PubMed=10660136;		
RA	Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,		
RA	Garcia-de la Guardia R., Urra S., Venegas A.;		
RA	"Serological response to Helicobacter pylori recombinant antigens in		
RT	Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and		
RT	gastric cancer."		
RL	APMIS 107:1069-1078 (1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CH-CTX1;		
RC	Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,		
RA	Bruce E., Mancilla M., Valenzuela P., Yudelevich A., Venegas A.;		
RT	"Cloning and comparison of ten gene sequences of a Chilean		
RT	Helicobacter pylori strain with other Helicobacter pylori strains		
RT	revealed higher amino acid sequence variability for VacA and CagA		
RT	virulence factors."		

us-09-077-574a-2.rspt

Mon Feb 2 12:04:10 2004

Search completed: January 28, 2004, 13:17:17
Job time : 42 secs